

Cover Page

Cyberinfrastructural Support for the Genetic and Ecophysiological Decipherment of Plant Phenological Control in Complex and Changing Environments

Grand Challenge Team Principal Members

Justin Borevitz (Asst. Professor, Department of Ecology and Evolution, University of Chicago, Tel 773-702-5948, Email borevitz@uchicago.edu). **Research & Outreach:** How do organisms shape their environment, and how does environmental change shape the resident organisms? The Borevitz laboratory is pursuing the adaptive genetic events enabling ecological succession, in annuals (*Arabidopsis thaliana/lyrata*) and perennials (Aquilegia, prairie grasses). The Indiana Dunes National Lakeshore is our model ecosystem. Outreach includes teaching summer courses for high school bio-teachers and an undergraduate PrairieEcosystems.org course that works with community and school groups.

Edward Buckler (Res. Geneticist, USDA/ARS Plant, Soil, and Nutrition Research Unit, Cornell University, Tel 607-255-4520, Email esb33@cornell.edu). **Research & Outreach:** Interested in the basis of complex traits with an emphasis on dissecting abiotic stress tolerance in maize involving quantitative and statistical genetics, genomics, and field genetics. We also have projects ongoing in wild and domesticated grape and switchgrass. Outreach conducted through informal science exhibits.

Melanie Correll. (Asst. Professor, Dept. of Agricultural and Biological Engineering, Univ. Florida, Gainesville, FL 32611-0570. Email correllm@ufl.edu Phone: (352) 392-1864 Ext. 209. Fax: (352) 392-4092) **Research & Outreach:** Abiotic stress in plants, root physiology, root functional genomics, ecophysiological modeling. . Works with high school students in the Student Science Training Program (SSTP) and 4-H, undergraduate mentor in Science for Life Program at UF (HHMI).

Chad Myers (Asst. Professor, Dept. Computer Science & Engineering, Univ. of Minnesota Minneapolis, MN 55455. Email: cmyers@cs.umn.edu Phone: (612) 624-8306 Fax: (612) 625-0572) **Research & Outreach:** Integration and inference from heterogeneous genomic data, applying machine learning and data mining to understand gene function and model biological networks. Outreach efforts include interdisciplinary research opportunities for undergraduates from diverse backgrounds and a local summer bioinformatics education program for high school math, computer science and biology teachers.

Johanna Schmitt (Olney Professor of Natural History, Dir. Brown University Environmental Change Initiative, Dept. Ecology and Evolutionary Biology, Brown University, Providence, RI 02912. Email Johanna_Schmitt@brown.edu Phone: (401) 863-3435 Fax: 401-863-3839) **Research & Outreach:** Ecological genetics of plant responses to changing climate, particularly flowering time in *Arabidopsis thaliana* in natural seasonal environments Outreach includes extensive undergraduate research mentoring, service on the Regional Advisory Committee of the New England Plant Conservation Program, and developing a classroom module for Rhode Island high school teachers through Project ARISE: Advancing RI Science Education.

Stephen Welch. (Prof. Theoretical Plant Modeling., Dept. Agronomy, Kansas State Univ. Manhattan, KS. Tel: +1-785-532-7236. Fax: +1-785-532-6094. E-mail welchsm@ksu.edu). **Research & Education:** Systems simulation with emphasis on applications, genomics and crop modeling. Electronic information delivery and decision support systems, especially in agriculture.

Jeffrey W. White. (Plant Physiologist. USDA/ARS Arid Land Agriculture Research Center, Maricopa, AZ. Tel: +1-520-316-6368. FAX: +1-520-316-6330. E-mail Jeffrey.White@ars.usda.gov) **Research:** Using ecophysiological models and geo-spatial tools to study plant response to global change in agro-ecosystems. Outreach entails helping diverse stakeholders understand potential global change impacts.

Summary

The phenology of a plant constitutes the temporal framework within which morphology elaborates, the resultant organs either grow or senesce, metabolic processes occur, and, ultimately, the genotype's contribution to the next generation (its fitness) is made. The proper timing of life cycle events is critical to the avoidance of both biotic and abiotic stressors and to achieving synchrony with beneficial species such as pollinators or seed dispersal agents. However, phenology is strongly influenced by environments that are changing in uncertain ways under the impact of human activity. Empirical ecophysiological models of plant phenology under field conditions date back over 270 years, but molecular insights into the reasons for their accuracy exist only for flowering time in *Arabidopsis thaliana*. Understanding the control of plant phenology and (ultimately) being able to predict it across a broad range of species is a problem which, due to its complexity and the lash of accelerating environmental change, rises to the level of a **Grand Challenge**.

Specifically, answers are urgently needed to the following scientific questions:

1. What are the genetic and physiological factors controlling phenology in non-constant environments?
2. How do the multiple signaling pathways/subsystems into which these factors are organized integrate environmental information to produce observed phenotypes?
3. What are the ranges of genetic diversity regarding phenological events?
4. What selective factors influence plant phenology in the natural world, how do they adaptively modify phenological control mechanisms/networks, and how can we learn from this to enhance programs of artificial selection for these traits (and others)?
5. How can different model types (ecophysiological, molecular, statistical, etc.) be combined for tasks ranging from phenological prediction to gene discovery and network inference?

To address these questions a cyberinfrastructure (CI) is needed that supports:

1. Development, adaptation, and/or adoption of an ontology to unify plant data bases;
2. Tools for integrating metabolomic, transcriptomic, and current/future high-throughput data; this includes (i) construction of software links to diverse data types and (ii) development of an integrative Network Explorer to overlay and peruse the data;
3. Advancement of current text mining capabilities;
4. Incorporation of tools for generating gene- and network-level hypotheses from diverse data;
5. Integration of data on $G \times E$ interactions for species, landraces, and genotypes and facilitation of new technology for collecting such information;
6. Establishment of a resource for species-specific knowledge of metabolism;
7. Creation of quantitative models from the gene- and network-levels to ecophysiology;
8. Collaborative work, data sharing, and other forms of group activity.

To meet these needs, CI is proposed that comprises middle-ware that mediates among diverse databases and hardware resources on the backend and, on the front end, advanced analysis and visualization tools serving a variety of audiences from plant biologists and stakeholders to students, educators, and citizen scientists. The CI recognizes, exploits, and supports commonalities that will leverage, assist, add value to, and draw sustenance from CI's proposed by several other Grand Challenge Teams. Community buy-in is documented via 89 collaborators who have agreed to participate in CI development and evaluation.

Key collaborations will also address the CI's education and outreach goals:

1. Train users on CI use in research, education, and outreach, and on CI extensibility mechanisms.
2. Increase the participation of underrepresented groups in the ecological, evolutionary, and agricultural associated disciplines.
3. Promote the active participation of graduate students in education and outreach activities.
4. Educate the public, targeting K-12 students, on the effects of global climate change on plant phenology and stress-induced responses.
5. Provide stakeholders (e.g., farmers, plant breeders) with tools that quantitatively predict phenology and stress responses for crops in relevant, real-time environments.
6. Support citizen science monitoring of plant phenology and stress in natural and managed ecosystems.

Description of the Grand Challenge Problem/Question.

Introduction. Phenology, the timing of biological events such as development, growth, metabolism, reproduction, and senescence, is an essential component of individual fitness and crop yields, and can affect ecosystem fluxes. Phenological events such as bud burst are primary determinants of net productivity and influence numerous responses to abiotic and biotic factors. Proper timing of plant reproductive onset enhances the probability that offspring will experience favorable environments. This makes the trait of significant interest to students of plant adaptation. Phenology can additionally determine species range limits (1). It thus creates ecological convergences critical to community dynamics, which often depend on delicate synchronies between plants and other species (e.g., pollinators, competitors, and seed dispersal agents), whose timing is also environmentally variable.

However, environments are changing under the impact of human activities in ways that scientists are not yet able to predict with adequate certainty. These changes take place at scales from global warming to more localized alterations such as modified agricultural practices that may affect floral timing and other phenological stages via micro-environmental variables like light quantity or quality, temperature, and relative humidity. Changes in plant physiological processes also feed back to landscape-scale environmental changes in ways that are yet properly accounted for. For example, whereas single-generation experiments in elevated carbon dioxide have shown fertilization effects, studies over multiple generations suggest that CO₂ can lead to selection for earlier flowering and reduced biomass in as few as five generations (2). If such effects are widespread, they might reduce carbon sequestration. Current climate and geochemical simulators incorporate submodels of land surface processes that include plant physiology (3), but not such evolutionary responses that operate on contemporary time scales.

Ecophysiological models simulate the developmental phenology as a accumulation of small, successive increments. The magnitude of each increment is calculated from dynamic environmental conditions at its corresponding point in time (4-12). Developmental transitions are predicted to occur when running totals reach threshold values determined from data. The molecular genetic explanations for the often high accuracy of these models are unknown, except partially for flowering time in *Arabidopsis thaliana* (10, 13). The long term aims of this proposed Grand Challenge are to (i) enhance quantitative prediction of *A. thaliana* phenology, (ii) extend these studies to flowering in selected other plant species of agricultural and ecological importance, (iii) leverage the results to create gene-based models for other phenological traits in those species, and ultimately (iv) throughout the plant kingdom. Significant progress on the first three steps is expected within 18-24 months, and methods will be well-established within three years, with ongoing progress toward others over the next decade. The outcomes will impact fields ranging from plant breeding to conservation of natural plants as affected by climate change.

The flowering time example. Flowering time has been intensively studied by developmental geneticists, evolutionary ecologists, and crop scientists. Although it is highly heritable and is often readily predicted for specific varieties in agricultural settings (4-12), there is still great uncertainty in our understanding of its control and hence, our ability to predict flowering dates of different genotypes in new environments. However, the time is ripe for an interdisciplinary synthesis to solve this problem.

The underlying molecular mechanisms are rapidly being elucidated. Over 100 genes are known to influence flowering in the model plant *Arabidopsis thaliana* (Fig. 1). Environmental information flows through (i) the photoperiodic promotion pathway that senses inductive long day lengths; (ii) the vernalization promotion pathway that responds to extended cold by down-regulating *FLOWERING LOCUS C (FLC)*, a major repressive integrator gene; (iii) the gibberellin pathway that is stimulated by levels of growth regulators; and (iv) a set of autonomous genes that rout promotive internal and external cues in parallel through *FLC*, leading to flowering independently of other factors (14-19). Other environmental factors such as shading by competitors (20) and drought can accelerate flowering via mechanisms that are less well understood.

The threshold values representing developmental transitions in phenology models may equate to gene products escalating to concentrations that activate molecular switches (21-23). The two genes *LEAFY (LFY)* and *APETALA1 (API)* exhibit a number of switch-like traits: (i) they mediate between environmental signal integrators and the floral organ identity genes (Fig. 1); (ii) they exhibit mutual positive feedback (a switch action prerequisite), possibly operating through *TFL1*; and (iii) their expression levels rise during floral induction (24-28). The network motif in which external or internal

signals trigger a switch which activates an entire downstream development subsystem has been observed in other organisms (29) and may well apply to other plant stages.

Yet, even for the one trait of flowering time, this wealth of data on one species leaves much to be discovered if we are to address the issues identified above. For many crops, photoperiod and vernalization loci are known, but there is less certainty regarding traits such as earliness per se, and knowledge of network structure is still far from complete.

Comparisons with *A. thaliana* suggest there are cases both of

homology and of independent evolution of mechanisms (e.g., for vernalization in wheat). Even when attention is restricted to the two inputs of temperature and photoperiod, there is enormous variation within the flowering plants. Salisbury (30) recognized 770 distinct photothermal floral response categories. But his system was still incomplete, omitting, for example, ambiphotoperiodic plants that flower in short- or long- but not in intermediate day lengths (31). He also misclassified *A. thaliana* as day-neutral, likely due to the great variation that exists within this one species. Adding other phenological traits, some of which may be quantitatively more responsive to other cues (e.g., carbon status; 32) exacerbates complexity.

Quite recently, it has been discovered in *A. thaliana* that there is a narrow (ca. 20-day) window in the late summer where a few days' change in germination time can alter flowering time by 100 days or more (13). Because of its sensitive, genotype-specific nature, this window plausibly has significant selective consequences. The sensitivity varies geographically, becoming acute at high latitudes, so any selective effects will likely synergize with climate changes that are also extreme in these regions. Moreover, mathematical experiments (Welch, unpub.) show that the windows are not photothermal effects per se, but intrinsically associate with the slowing (but not acceleration) of developmental rates in response to any relevant, seasonal variables. In many areas of the world, seasonal variation is not photothermal in character, but rather a wet-dry cycle, and water stresses (both drought and flooding) can influence flowering time. (Note to reviewers: An independent companion proposal focuses on the effects of several stressors, including water, on plant processes other than phenology.) Thus, such windows are probably present in most flowering plants to degrees and at times that depend on local conditions.

THUS, UNDERSTANDING THE CONTROL OF PLANT PHENOLOGY AND PREDICTING IT FOR MANY SPECIES IS DIFFICULT, COMPLEX AND, GIVEN THE URGENCY OF ACCELERATING ENVIRONMENTAL CHANGE, RISES TO THE LEVEL OF A **GRAND CHALLENGE**. The specific scientific questions to be addressed are:

1. What are the genetic and physiological factors controlling phenology in non-constant environments?
2. How do the multiple signaling pathways/subsystems into which these factors are organized integrate environmental information to produce observed phenotypes?
3. What are the levels of genetic diversity for phenological events in different species or ecotypes?

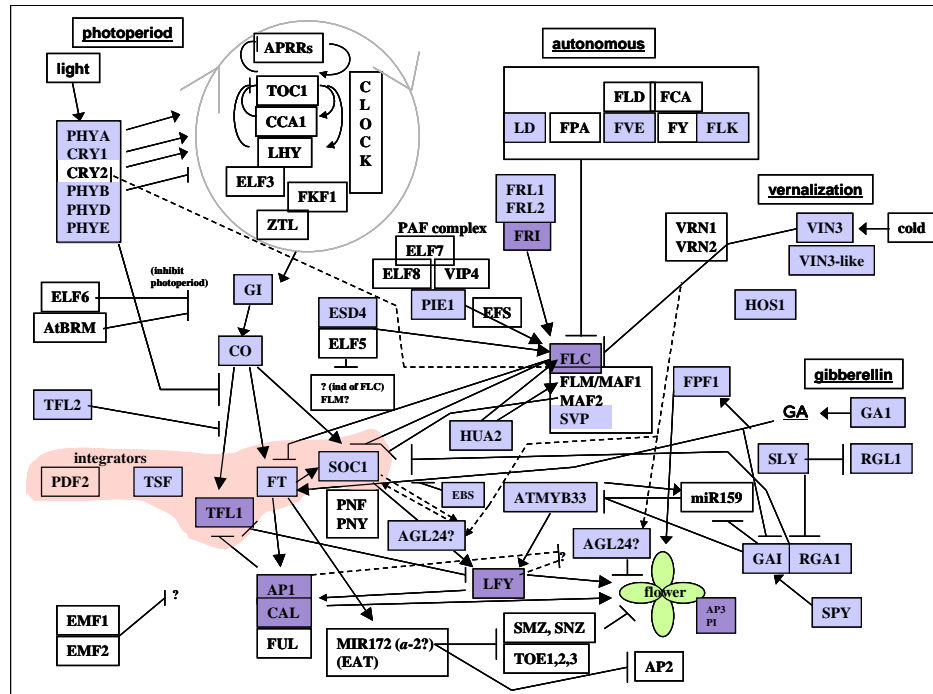


Figure 1. A subset of the Arabidopsis flowering time genetic control network. The color codes are not significant in this context. (Diagram by JL Roe)

4. What selective factors influence plant phenology in the natural world, how do they adaptively modify phenological control mechanisms/networks, and how can we learn from this to enhance programs of artificial selection for these traits (and others)?

The requirements necessary to meet this challenge are elaborated below. One, however, deserves early mention – a sophisticated quantitative modeling capability. Empirical ecophysiological models of flowering time based on the thermal summation concept date to 1735 (33), and are quite skillful in field settings. Gene-based phenological prediction in crop species shows promise, but models are still rudimentary (e.g., 34-36). Molecular systems biology models of the diurnal clock (e.g. 37, 38), a critical component of flowering time control, are quite advanced. And progress has been made in relating the success of ecophysiological flowering time models to molecular events (10, 13, 39). Statistical models are used for association mapping of phenology and many other traits. A methodological question is:

5. How can these different classes of models be combined or otherwise used together for tasks ranging from phenotype prediction to gene discovery and network inference?

What impact would a successful infrastructure development would have on the broader field?

Several of the scientific questions above have implications well beyond phenology. For example, the “non-constant environments” in Question 1 can readily include not only current conditions but future climates (40). Knowing the factors controlling particular phenological events will enable georeferenced database searches of abiotic conditions at global, regional, and local scales (e.g., see Table A-1) for sites defining species-specific “climate windows”. These windows could be realistically narrowed using data from similar searches targeting competing plants and (albeit out-of-scope for this project) phytophagous and/or pathogenic species. Question 2 is an instance of the genotype-to-phenotype (G2P) problem, arguably the most important problem in applied biology today (41). In fact, a recent report of the National Research Council recommended that “...the National Plant Genome Initiative should have as its top priority ... science aimed at a detailed, system-wide understanding of the functions of single genes, how those functions are connected in networks, *and how they control plant growth, form, function, and evolution*” (emphasis added, 42). Phenology (especially flowering time) is an ideal model trait for developing G2P methodologies because, while much remains to be learned, there is enough information to evaluate new ideas. Question 4 specifically refers to learning new knowledge from phenology to improve our capacity to deal with other traits.

Question 5 applies far beyond phenology (e.g., it is of particular interest to the C4 iPlant proposal team; Brutnell, pers. comm.). Quantitative models are being developed in many branches of plant science (43-48). The NRC report just cited also recommends the use of “... system-level approaches to understand plant growth and development in controlled and relevant environments to create the iPlant, a large family of mathematical models that generate computable plants genuinely predictive of plant system behavior under a range of environmental conditions”. However, for the potential power of modeling approaches to be fully realized, differing classes of models must (i) synergize both as software *and* as mathematics, while (ii) exploiting *and* delivering information about processes and organisms from/to disparate datasets and scientists in multiple disciplines.

The commonality of these needs suggest that cyberinfrastructure (CI) developed for research on phenology will benefit a much wider range of plant scientists. As examples, the CI proposed here would aid plant stress studies described in a companion proposal and elements of the CI’s in the Tree-of-Life, C4 photosynthesis, and BIEN submissions would be useful in phenology research.

What scientific activities would a collaborating team undertake to make progress solving the GC?

Figure 2 places phenology studies in the context of a wide range of plant biology research that the proposed CI will also aid. Exploiting large amounts of data on phenotypes (including mutant), environmental dynamics, “omics”, sequence information, association mapping, and other analyses, collaborating teams would identify genes of interest and their variability in either natural evolving systems or in sets of germplasm under artificial selection (Questions 1 and 3). While studies of *A. thaliana* floral initiation will continue, expanding efforts would address other post- and pre- flowering traits (e.g., first real flower, seed development in relation to flowering, bud set and bud burst in woody perennials, phyllochron intervals, and tuber initiation). The power to detect new genes would be

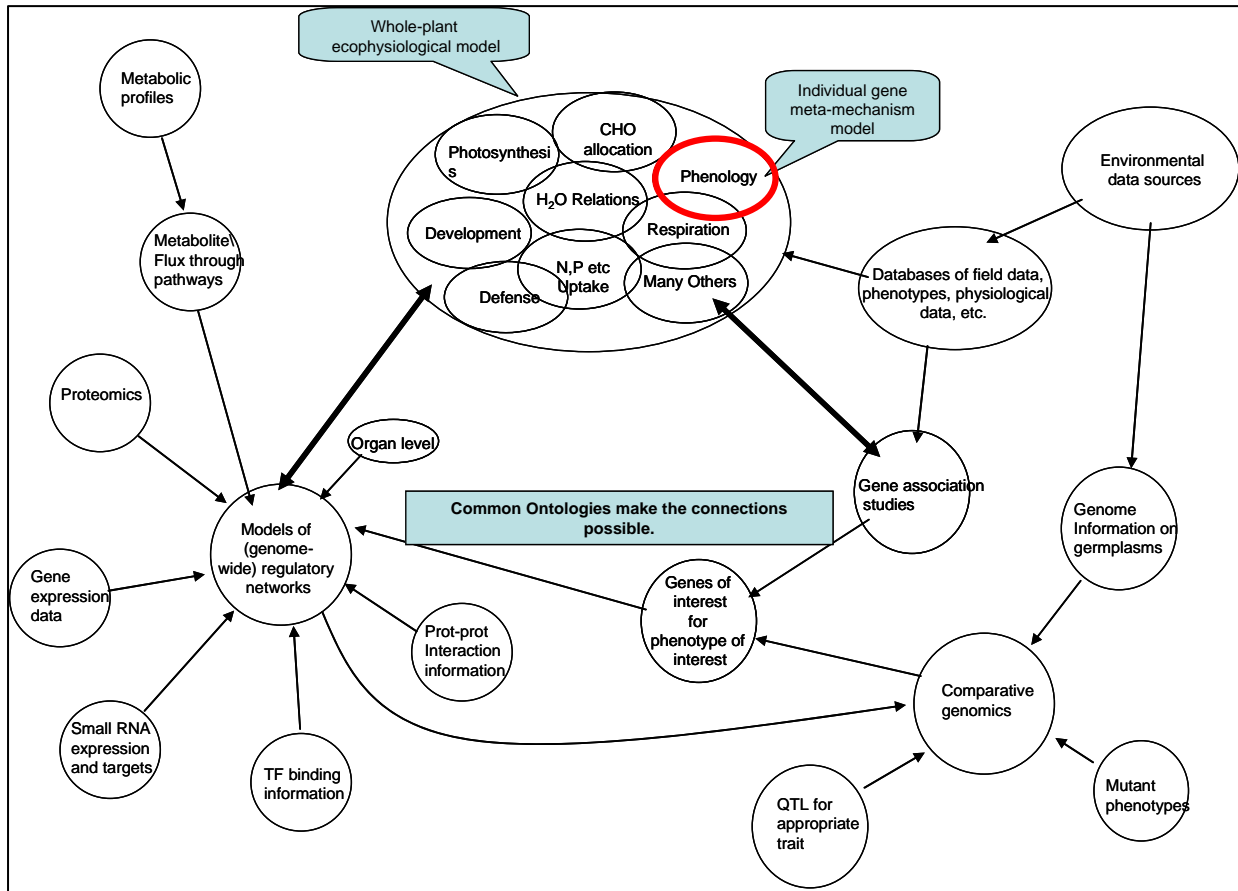


Figure 2. Differing communities work in parallel on various parts of the Grand Challenge using data sets, analysis tools and models in the process. New results are stored in databases and/or used to construct/improve useful links. In turn, the revised models may suggest further experimentation and/or provide predictions of broad use within, and externally to, the project. (Picture by K. Denby et al.)

enhanced in novel ways via the ability of progressively gene network-based ecophysiological models (Question 5) to explain *GxE* effects (right bold black arrow; 49-52).

Concurrently, research to infer network structures (left part of figure, contributing to Question 2) would combine a broad spectrum of high-throughput, high-volume molecular data. The CI comprises an analysis pipeline of automated data acquisition, supervised integration of current and future data on gene expression, metabolic pathways, protein interactions, molecular interactions, regulatory interactions, and cell-specific expression. Initial studies would focus on plants for which significant information is available such as *Arabidopsis*, rice, wheat, barley, maize, sorghum, grain legumes (soybean, common bean, pea, and peanut), and *Mimulus*, but expand over time. As collaboration builds, vocabularies will become more commonly understood and standardized via ontologies, enabling computational linkages to be made and strengthened throughout the system.

Gene-based, whole-plant ecophysiological models would provide higher levels of synthesis (left bold black arrow). Such models, under development for over 40 years (53), combine physiological principles with empirical mathematics. They predict plant responses (phenology; dry matter production and partitioning; and growth) under non-constant conditions of temperature, solar radiation, water, and nutrients. For flowering time, alignment has been found between these empiricisms and the underlying genetic mechanisms (13). When present (or constructed) such alignments (which may be called *gene metamechanisms*) quantitatively model the overall, summary behaviors of the underlying networks in simplified form. Simulations and sensitivity analyses of these models (13) in relevant populations of environments would provide estimates of fitness and selection differentials pertinent to Question 4.

What are the current barriers to performing these scientific activities?

Researchers currently lack an integrated view of QTL, “omics” data, sequences and markers, genes, proteins, phenotypes and environmental measurements (georeferenced and time/date stamped in the case of field observations), and simulator outputs detailing possible future environments. Although some data are now managed through relational databases, such as TAIR, PlantGDB, ICIS implementations, and GRIN, many important datasets still reside in individual publications, spreadsheets, or printed reports with minimal metadata or quality control. Even when data are in accessible digital formats, researchers from slightly different fields may have difficulties interpreting reported data.

Another barrier is the vocabularies of the participating research communities. Ontologies, such as those being developed for genes (www.geneontology.org) and plant traits (www.plantontology.org), can lower these barriers, but to date, lack of software tools limits use of ontologies in the phenology community. Tools must be built to promote use and enhancement of genomic, phenotype, and environmental ontologies in conjunction with established ontology groups such as POG.

Many available models lack effective interfaces. These include tools for preparing and checking model inputs and for visualizing multidimensional sets of quantitative outputs. Similarly, better tools for parameter estimation and for evaluation of outputs would let scientists focus on improving models and applications. Decision support systems linked to crop models, such as the DSSAT software (54), contain tools that might be adapted or serve as prototypes for components of the integrated system.

Mechanisms are needed to aggregate information from molecular and cellular scales and to transfer it to higher biological levels (e.g., leaf or canopy). Then, by exploiting clues from orthology, synteny, network resemblances, similar forms of ecophysiological behavior, or other evidence, we need to be able to make inferences across species. Only such tools will let us solve the genotype to phenotype problem. Few plant biologists can accomplish such syntheses alone. Some graduate programs are providing the requisite training for the next generation (see below). The work proposed here is well -suited for interdisciplinary teams of graduate students, post-doctoral associates, and faculty. In the longer term, the CI could provide live-fire exercises (i.e., with real data), and study materials to increase user proficiency.

Description of the computational abilities and (available) data and needed for these activities.

Discussions between attendees of the Climate Change, Plant Adaptation, and Mechanistic Modeling Grand Challenge Workshops identified major commonalities in the core features for the desired CI. The CI should be extensible as to data sets, analysis programs, and other available features; and provide interoperability among its many components. These traits will let the system stay current with advances in measurement and analysis, support individualized development by particular communities, yet allow capabilities with general utility to diffuse rapidly to broader audiences.

FOR THIS REASON, THE CI PROPOSED HERE IS IDENTICAL TO THAT SUBMITTED BY THE STRESS GROUP.

The developmental time required for each of the following needed items is indicated (long-term, LT; intermediate term, IT; short-term, ST; or prototyped in the short term, PST).

- A. *Develop/adapt/adopt an ontology to unify plant databases.* The use of a controlled vocabulary is a key component of database interoperability. It should span/contain information on environments (laboratory, field, or other); taxa; phenotypes at multiple organizational levels; associated genes; their roles in cells; and/or other features of relevance. This activity is ST relative to any one project, but should be an on-going iPlant effort, cross-cutting all current and future GC Teams.
- B. *Develop data integration tools for integrating (i – ST) metabolomic and transcriptomic data and (ii – IT) current and future high-throughput data.* There are two aspects to this latter task. (1) The first is to identify sources of sequence, transcriptomics (e.g. from different platforms posted at GEO across/within crop species and Arabidopsis, ChIP-Seq, and DNA binding sites), proteomic, interaction (physical and genetic), and phenotype data. Several databases where such data are available are listed along with other data sets in Appendix Table A-1 (the curators of some of these databases are collaborating in this project). The CI will not necessarily need to directly involve storage of all data, instead it will provide a common interface to talk with other genomic databases (e.g. NCBI gene expression omnibus, bioGRID). A good model for this is the Yeast Functional Genomics Database (<http://yfgdb.princeton.edu/>), which allows the yeast community to access almost any type of published yeast data. (2) The second is to develop an interactive user interface for

integrated browsing of high-throughput data and computationally-derived metadata. This web-based interface should support the overlay of various experimental or curated data and enable exploration of computationally derived network and gene-level hypotheses. It should be possible to visualize selected parts of the network, showing connections supported by various types of data and inference methods (see D). Nodes should be collapsible into “super nodes” to reveal interactions between larger scale phenomena. VirtualPlant (www.virtualPlant.org) uses tools to integrate visualization of network and other data, including SunGear (<http://bioinformatics.oxfordjournals.org/cgi/content/full/23/2/259>) and Cytoscape (55). This activity would benefit greatly from close links to the Tree-of-Life project.

- C. *Develop text-mining tools.* This consists of three integrated activities: (i – ST) Select one or more extant tools (<http://genomebiology.com/supplements/9/S2>, <http://bionlp.sourceforge.net/>) to integrate with the system. These tools can determine that a gene/protein is mentioned in an abstract and detect interactions. As with other data management features, it is the integration that adds value. (ii – LT) iPlant will work in concert with GC Teams and publishers to embed markups in electronically published editions to facilitate automatic information extraction technologies. Expanded text mining by our TAIR collaborator is now under NSF review. (iii – IT) Graphs and tables in scientific papers often contain crucial information that is not accessible to text mining tools. Until that information is made accessible using embedded markup, we propose to create a tool that recognizes tables and graphs appearing in a paper of interest, and with user feedback, extracts the data into files with common formats (e.g., spreadsheets). For examples of prior work on table extraction see (56, 57). Verification mechanisms to aid human comparison of original and extracted data will be included. Expanded text mining by our TAIR collaborator is now under NSF review.
- D. *Compare/adapt/adopt computational data integration/fusion tools for generating gene-level and network-level hypotheses from diverse data.* We propose that iPlant staff adapt existing approaches for integration and learning from diverse genomic data, including sequence, co-expression, epigenomics, proteomic, metabolomic, and protein-protein interaction data, and data mined from the literature. Recent work in this area has focused on applying machine learning to predict functions for uncharacterized genes (58-61) and to suggest “functional linkage networks” (62), which capture general associations between proteins implicated in the same biological processes. Both of these approaches are effective strategies for identifying new candidate genes. We anticipate that static network models based on an integration of a wide variety of data will prove useful as a starting point for more mechanistic modeling efforts, particularly for sub-systems that are less well understood. Examples of existing systems that could be adapted for these purposes are GeneMANIA (63), STRING (58), and bioPIXIE (62). More specialized methods for network inference, including regulatory network inference methods (64-66) or boolean network models (67) based on time-series data could also be incorporated as part of this component of the infrastructure, as well as tools for cross-species network alignment (68, 69).
- E. *Integrate data for plant species, landraces, and genotypes.* There are large sets of existing data on rice, maize and sorghum diversity and other crops (see Table A-1) that document specific genotype/environment interactions for species, landraces, and genotypes. Work under this topic will involve collaborating with the curators of such data sets to incorporate them into the CI. (ST) Also, technologies of potential relevance to field data collection continue to advance and now include everything from barcode readers to cell phones with imaging and GPS capabilities. Complete pipelines are needed that reach from such devices to CI data sets. (PST – IT)
- F. *Establish a resource for species-specific knowledge of metabolism,* in collaboration with the curators of PlantCyc, that combines orthology mapping such as in MetaCyc plus literature information, cross referenced with other databases and machine-readable structures in InChI codes. After curation (i.e. excluding agrochemicals), such new or combined and cleaned repositories could then serve as entry point for adding physiology, biochemical and experimental evidence data for model building.
- G. *Develop quantitative models* from network structures (e.g., gene regulatory, metabolic, or ecophysiological process interactions) produced by inference tools or user-entered. This involves (i) associating mathematical formulas with network links; (ii) visualizing model behavior; (iii) estimating parameters; (iv) performing sensitivity analysis; (v) juxtaposing model outputs graphically with data, and performing statistical tests or other tasks from QTL mapping to climate effect research; (vi) linking multiple models within the CI; (vii) driving models with inputs from CI databases; and

(viii) storing outputs for later use. Tasks (ii to viii) should be applicable to statistical, L-systems, and ecophysiological models, alone or at levels of integration nearing virtual fusion. Models and data should be locally storable or publishable to the broader community. Creating prototype models, to suggest needed experiments or for scientifically important, unmodeled species is a ST or IT task; model refinement, inter-linking and other forms of model integration will extend into the LT.

H. *Provide groupware support* for team activities and for the above functions as they come online. The CI should support document sharing/management, group scheduling, task management, threaded interaction, text chat, audio/videoconferencing, and common desktops for joint software operation.

Because this capability will be central during both CI development and use, its priority is *Immediate*.

Project Principals will lend their expertise along with iPlant staff and interested community members to integrate and manage A through H. If the Plant Stress proposal is accepted, both projects would interact closely, with some differential emphasis. Both projects would contribute equally to A, C, D, H; this project would work more heavily in E and G; and Plant Stress would have a focus in B and F.

Considerations of comparative advantage suggest that all or part of certain critically important activities/developments should be outsourced by iPlant to team members working together with all interested others. These special responsibilities are presented in the Grand Challenge Team section.

Collaboration needs with computational researchers and available expertise on the GC team.

The GC Teams of the two companion proposals include plant biology and computer scientists, and computational biologists, including quantitative modelers. What will be needed is expertise in data and text mining, middle-ware, image processing (see the proposed text mining tool), software engineering, code writing, and designers for web-based services and outreach applications.

Projected cyberinfrastructure software and hardware needs, and potential bottlenecks.

Key early architectural design decisions concern the degree of (i) centralization to be employed in system implementation and (ii) reliance to be placed on existing systems and prototypes for core software. The range of alternatives for each point is continuous, but we believe that past history, the user adoption dynamics, and the economics of sustainability argue for relative decentralization and, to the extent possible, exploitation and expansion of existing systems rather than completely *de novo* creation.

The salient features of today's reality are that the aggregate (i) pace of innovation and the resources available to sustain it are far greater when summed across any community than they can ever be within a single, soft-funded unit; and (ii) 50-75% of the total life-cycle costs of software accrue after release in the form of maintenance, upgrades (<http://fast.faa.gov/pricing/c1919-6.htm>), and user support. Systems designed to be extensible by high-end users permit and pay for improvements in a distributed manner. Relevance is maintained via exponentially increasing user-driven/supplied capabilities. The alternative model of a soft-funded unit can lead to a dead-end unless the relevant community takes on the system maintenance costs once the original funding runs out.

It is important to note that this accelerating innovation will be ongoing even while the CI is under development. A rapid prototyping approach, wherein early (but not necessarily later) "releases" are based on extant systems, allows users to perceive value and even extend services from the beginning. Thus the system can maintain currency with their evolving expectations and expanding use while various aspects of development proceed in parallel.

Two adverse examples are illustrative. Stein (70) notes the slow adoption and lack of grass roots support for the caBIG system due in to from top-down management and extensive innovations in core features that are not, however visible to end users. In the early 1980's the WK Kellogg Foundation funded a multi-million dollar effort to network and microcomputerize County Extension offices in Indiana. Again the focus was on innovation in core features by a highly competent team, which, however, delayed delivery of end-user services. Users found this frustrating given the rapid market proliferation of microcomputer software that was unavailable to them due to system incompatibilities. In the end, the effort was not deemed successful.

To facilitate the adoption of the CI among tool and model developers we will hold data analysis competitions on tasks such as prediction of protein interaction networks, pathway prediction, prediction of phenotype responses etc. Such competitions can have several effects: (i) researchers will realize that the data integration infrastructure allows them to analyze data much more effectively; (ii) that making their algorithms available through the discovery environment is easier than having to design and

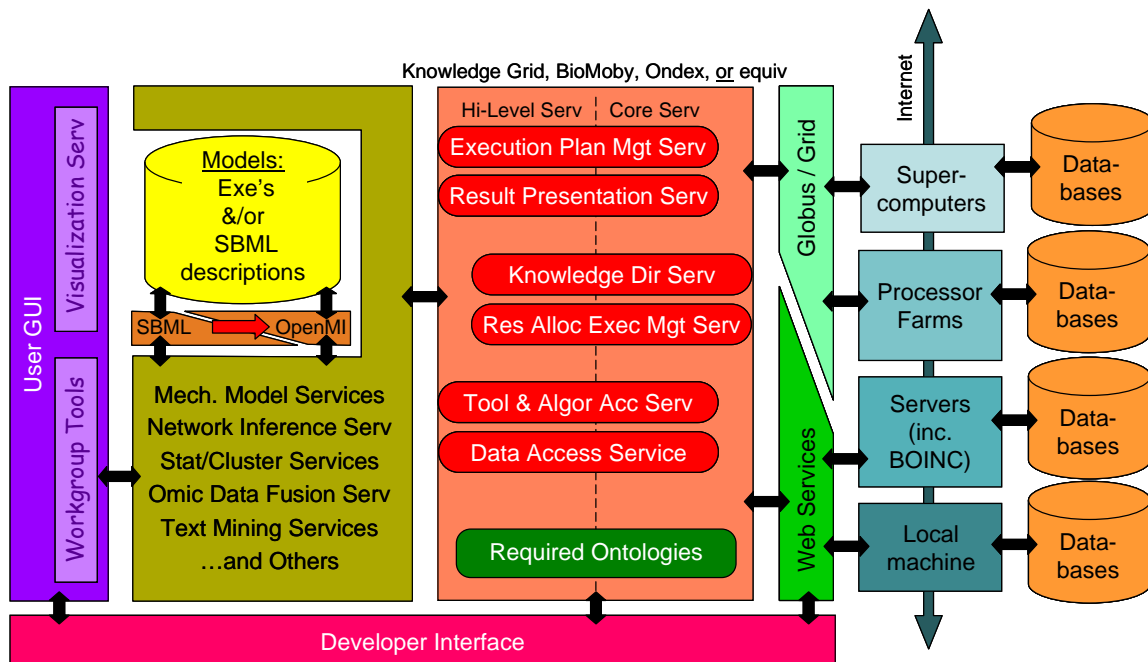


Figure 3. Computational operating software view of the CI. Individual blocks are described in the text. (Diagram by SM Welch et al.)

implement their own web-frameworks (*iii*) that they are able to reach a bigger audience with their tools. The end-result can be to attract more researchers to work on plant genomics/bioinformatics, thereby pushing the state-of-the-art in the field. Such competitions are extremely useful for the community since by having everyone analyze the same dataset, it is easier to figure out the strengths of the various methods; even once it is over, the resulting data remains as a resource for the community that later algorithms can be tested on. A recent example is the MouseFunc protein function prediction competition (<http://genomebiology.com/2008/9/S1/S2>). A member of the stress team (Ben-Hur) has co-organized such a competition (<http://www.clopinet.com/isabelle/Projects/NIPS2003/>) in the context of feature selection for discriminative classification.

If the system is to be extensible (Figure 3), it must possess some core entity that manages the addition process and makes new features available to users. Examples containing potential prototype cores (red central block in Figure 3) include, but are not limited to, the Knowledge Grid (71-73), BioMoby (see <http://biomoby.open-bio.org/index.php/what-is-moby/>), and Ondex (www.Ondex.org). Within their respective (and non-disjoint) spheres, these systems allow users to automatically locate data and analytical services and flexibly combine them toward various ends. The Knowledge Grid is used in what follows for illustration, but no endorsement is implied.

User Interfaces. The *Developer interface* is for users with sufficient expertise to add or modify system content, and the *User GUI* supports individuals querying/analyzing data and/or modeling. Developers range from software engineers creating or maintaining the system to researchers/algorithm developers adding new tools. The Developer interface permits entry, editing, and system debugging, creation or modification of system features, and publication of the results to the entire community.

The *User GUI* is for general system interaction. Figure 4 shows how elements of the system appear to users. Users are allowed to search for or enter data, or analyze/model/visualize them via pre-defined or newly created workflows. Aside from the code needed for the GUI and network interaction, CI resources need not reside on the user's local machine. Nor need users be aware of where data or execution resides, although it may be useful to retain this information for training or troubleshooting purposes.

System Hardware Level. Processor requirements range from single workstations to high performance supercomputers (HPC). For example, a mapping model with 40 QTL (ca. 1000) parameters requires three to four hours on a quad-core workstation using data from 5000 lines (PI Buckler). Using expression data would involve over 2×10^4 as many arithmetic operations. A multi-crop parameter estimation study (74;

PI Welch) required the ecophysiological simulation of seasonal plant growth and development for 3.8×10^9 site-year combinations. BOINC technology (<http://boinc.berkeley.edu/>) allows the general public to participate/contribute to numerically intensive computation.

Network protocols. Standard protocol stacks (Web Services and Globus) can be used to manage these heterogeneous resources, assign tasks, and handle inter-process and inter-processor communications. Versions of Knowledge Grid exist for both environments.

Interoperability middle-ware. Knowledge Grid (Figure 3; used for illustration) has a set of lower Core Services (right justified red ovals) and, building upon them, a group of High-level Services (left justified ovals). Services (71) included in these two levels are:

1. Knowledge Directory Service – searchable sets of XML documents describing published data sources; tools and algorithms for data management, analysis and visualization, and storage of results; and workflow plans. This metadata provides the basis for automatically finding needed data and tools and making them interoperable. (See box at right.) To maximize search effectiveness, these documents should be curated and utilize ontologies.
2. Resource Allocation and Execution Management Services – pair up the tasks in an execution plan with available resources (computation, storage, network bandwidth, and latency).
3. Data Access Services – find, extract, transform, and deliver required data.
4. Tools & Algorithms Access Services – locate needed tools and algorithms.
5. Execution Plan Management Services – semi-automatically generate one or more execution plans that interweave tools and data to meet user, data, and algorithmic requirements and constraints.
6. Results Presentation Services – specifies the ways to store, present, and visualize computed results.
7. End-User Services – produce results for the general user. These services are invoked by a workflow plan and may request run-specific information from the user. Outputs may be visualized directly or stored and further analysed/displayed later. The End-User Services in Figure 3 may be added to later using the CI's ability to be extended. Selected End-User services are:
 1. Visualization Services. These can be divided into several categories. *Network diagrams* include gene and metabolic networks (undirected, directed, and quantitative models), workflow diagrams, etc.

“Combining” Databases
 Consider an analogy in which *databases* are radio telescopes. Dishes exist all over the world, bigger scopes can see farther, but there are engineering (*i.e.*, resource) limits to how large any single one can be. *Metadata* is information about a data source. So metadata about radio telescopes might be size, location, etc. Using metadata, computer software (which, this analogy, equates to *middle-ware*) and networks, radio astronomers can meld all the distributed scopes together into a *virtual radio telescope* with a diameter equal to that of the Earth. This virtual instrument can see very well indeed, as will researchers using the databases, tools, and models we propose to similarly meld.

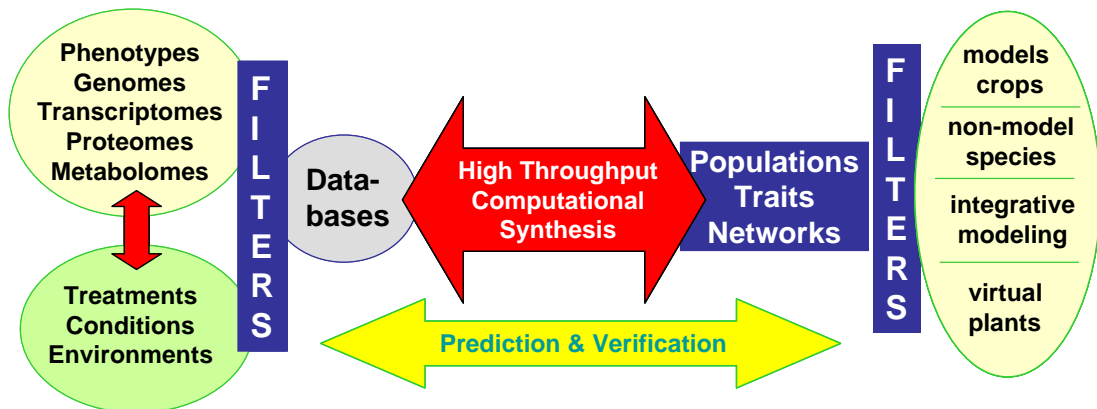


Figure 4. The Users' view of the system. (Diagram by members of the Stress & Climate Change Workshop organizing team)

Scientific data graphics includes conventional scientific graphs of data. *Genome mapping* includes sequence viewing, alignment, annotation, QTL graphing, etc. *Geographic mapping* includes the display of georeferenced data, which are critical for climate change and adaptation studies.

2. Mechanistic Model Services. Networks can be modeled in Systems Biology Markup Language (SBML; www.sbml.org/Main_Page), which captures both qualitative features (e.g. connectivity) and quantitative equations in Math Markup Language (MathML; <http://www.w3.org/Math/>). Many tools support work with SBML models. Ecophysiological models exist (as executable files) for all major crops, many minor ones, and even weeds of economic significance and are important resources for solving the genotype to phenotype problem (75). Finally, plant morphological dynamics can be modeled by L-systems (<http://algorithmicbotany.org/FSPM07/proceedings.html>; 76).

These different types of models need communicate with each other and with other CI services. SBML embodies sufficient mathematical expressivity that, in principle, it can fully describe an ecophysiological model, however experimentation is needed to see how fully practical this is. Even so, the wealth of SBML tools means that it will be useful to express certain ecophysiological sub-process models in SBML. Alternatively, executables can interact via OpenMI, originated by EU hydrologists and gaining wide use as a model interface standard (<http://www.openmi.org/reloaded/>). Models can be OpenMI compliant *ab initio* or be made so by surrounding legacy code with a wrapper. Putting such wrappers around SBML tools (red block arrow in Figure 2) and L-system interpreters would allow all three model types to communicate.

3. “Omics” Data Fusion Services –The desired services will exploit the Knowledge Grid metadata to search, as needed, for relevant databases and mine them for relevant information. It may either (i) utilize the Visualization Services to present the results in user friendly form, (ii) create and deliver to the requestor for subsequent use a specialized database containing the requested data in integrated form, or (iii) create such a database and publish it to the community (i.e., generate the required metadata for inclusion in the Knowledge Grid).
4. Statistics and Clustering Services – Handle a wide range of tasks from routine data analysis to QTL mapping to data dimensionality reduction. Two example candidates to serve in this or a supporting role include *R* (<http://cran.r-project.org/>) and TASSEL (77). An example of a grid-based QTL mapping system is http://www.gridqtl.org.uk/gridqtl_project.htm.

What falls within the scope of the GC Collaborative Project and what is considered out-of-scope.

We propose to use current and future data to drive the construction of conceptual and quantitative regulatory/ecophysiological models for plant phenological responses to complex natural environments, including those to be expected under climate change. Although climate change factors are the motivation that drives this proposal and the work that is planned, linking the models that arise from the work described here to climate change models themselves is beyond the scope of the GC envisaged.

Education and Outreach

A critical need exists to prepare the existing and future workforce to address pressing problems such as global food supplies, natural resource conservation, and climate change. Many aspects of plant science are relevant to these much-debated issues. The common CI proposed here and by the Stress group can support the education and outreach efforts of both teams with particular examples contributed by each. This synergy is further promoted by the presence of Correll on both GC Teams. Our efforts are designed to add value to currently funded programs. The main goals for education and outreach are:

1. Train users on CI use in research, education, and outreach; and CI extensibility mechanisms.
2. Increase the participation of underrepresented groups in the ecological, evolutionary, and agricultural associated disciplines.
3. Promote the active participation of graduate students in education and outreach activities.
4. Educate the public, targeting K-12 students, on the effects of global climate change on plant phenology and stress-induced responses.
5. Provide stakeholders (e.g., farmers, plant breeders) with tools that quantitatively predict phenology and stress responses for crops in relevant, real-time environments.
6. Support citizen science monitoring of plant phenology and stress in natural and managed ecosystems.

Workshops, learning modules, and online educational tools (Goal 1). Flexible learning modules in CI operation/expansion will be needed for students, researchers, stakeholders, developers of system add-ons, etc. We will develop these modules with the iPlant Education and Outreach Team. This is an early priority to support the CI rapid prototyping necessary to fuel/sustain community interest and support. The preparation of instructional modules, itself, will be a training priority so as to enable rapid expansion of contributed materials, for example higher level exercises that include end-user research. Conferences and extension-outreach will be used as training venues and include, the annual tri-society meetings of the ASA-CSSA-SSSA, the ASPB, the AAAS and others. Online tutorials and downloadable classroom modules will allow teacher- and/or self-directed discovery.

Recruitment of underrepresented groups into Grand Challenge science (Goal 2). Most universities have a variety of outreach organizations. All of these venues can be enriched via the types of educational modules, models, and research capabilities the CI will supply. For example, PI Welch (Kansas State University) has interacted with two such groups that can serve as example contact points. (i) Students in the Developing Scholars Program (DSP) choose a faculty mentor based on their research interests and, together, design a research project. Since 2000, the DSP has tripled from 20 to 60 students. (ii) The Ecological Genomics Institute (EGI) of Kansas State University has programmatic connections with Black Hills State University (BHSU), which lacks a PhD program but has the only Ecological and Evolutionary Functional Genomics Masters program in the nation. An NSF-funded scholarship program at BHSU helps underrepresented and needy students, especially Native Americans, transition from undergraduate to Ph.D. and professional STEM programs with their MS program its focal point. A main goal of the KSU/BHSU partnership is to establish a “pathway” for student transition between the BHSU MS and the KSU PhD programs. Welch is an EGI Steering Committee member and thus positioned to infuse the products and resources offered by the CI into these and other EGI activities that recruit and sustain underrepresented groups.

In the companion stress project, Grene liaises for Virginia Tech in the Virginia-North Carolina Alliance for Minority Participation (VT-AMP; <http://www.maop.vt.edu/programs/vancamp.html>), funded by the Louis Stokes Alliance for Minority Participation at the NSF. The program promotes the graduation of minorities, and their advancement into the profession, in the STEM fields. The Alliance includes research institutions, such as VT and the University of Virginia, and several Historically Black Colleges and Universities (HBCUs) across the two states. Select first-year VT students in STEM majors are invited to participate in VT-AMP, beginning with the AMP College Transition Week prior to the start of fall classes. Transition Week provides VT-AMP students with structured academic programming to facilitate the high school-to-college transition. Transition Week includes a variety of workshops, including: Introduction to Research, and the Seminar Series on Academic Success. VT-AMP Scholars receive a stipend for participating in VT-AMP programs and completing a research project with a VT faculty mentor (required). Structured academic support continues through the academic year. With funding of this iPlant proposal, VT-AMP scholars will be recruited to work with members of the Grene group on the stress/climate change project.

Public education (Goals 3, 4). NSF encourages graduate student participation in education and outreach activities, rising, in the case of the Plant Genome Program, to their involvement as outreach/education coordinators. If both proposals are funded, all graduate students and post-doctoral associates will work collaboratively with the iPlant Education Team on the design and development of CI tools to educate the public on the science behind the Grand Challenge topics as well as CI use. Correll is working with the NSF-funded University of Florida program “*Science Partnership in Inquiry-based Collaborative Education*” (<http://www.spice.centers.ufl.edu/what.asp>). SPICE seeks to (i) increase graduate student communication skills and foster a long-term commitment to outreach activities, (ii) encourage underprivileged middle school students to pursue STEM careers by enhancing science education, and (iii) develop strong and enduring University/K-12 partnerships. The SPICE program funds graduate students to develop education and outreach modules for of 6th to 9th grade teachers on the topic “Ecosystem Health and Sustainability”. SPICE fellows spend at least two days a week in middle school classrooms. Collaborative work with SPICE program will (i) provide a test bed for educational tools developed in the CI, and (ii) promote the education of graduate students (SPICE fellows and iPlant-funded students) on effective ways to teach the public about the effects of global climate change on plant responses.

Stakeholder support (Goal 5). Because of the importance of phenology and stress responses of plants in crop management, farmers frequently seek guidance on expected time of flowering and maturity of their crops and how their crop will respond to different environments. The project will collaborate with ongoing work of the Southeast Climate Consortium and CLIMAS to improve the ability of the [AgroClimate.org](http://www.agroclimate.org) web site to predict phenology of crops relevant to their regions. For example, Michael Ottman, U. of Arizona extension agronomist, wants to replace his current email advisories for wheat and barley with on-line tools that allow growers to select specific locations, cultivars, and planting dates and receive predictions from models that incorporate the best available understanding of physiological and genetic control of phenology. The CI would provide tools for linking AgroClimate.org to the requisite cultivar data on loci known to affect phenology and model responses. The developed materials will be available on the iPlant website and link to other education and outreach sites (e.g., Plant Genome Research Outreach Portal; <http://www.plantgdb.org/PGROP/pgrop.php?app=pgrop>, e.g., Southwest Climate Change Network <http://www.southwestclimatechange.org/>).

Citizen science (Goals 4 and 6) The PI, Borevitz, is working with the National Phenological Network to organize field observations from around the country. (The support letter from NPN Exec. Dir. Jake Weltzin documents this as well as novel research opportunities.) NPN is a sponsoring organization of Budburst.org which is targeted toward citizen scientists and school groups. Individuals can follow the development of target plants and/or species in nearby locations. They can upload their observations to the community network. Species identification cards and help sheets are available at budburst.org. In practice however, this format is slow and limits users to few plants at a few sites.

Several new tools are available that leverage user-operated GPS capability. Geotagging allows users to sync GPS coordinates with digital images taken in the field. This can be achieved with compatible PDA phones or standard cameras and \$100 GPS tracksticks. This allows an individual to position hundreds to thousands of plants in a single day. This high spatial and temporal resolution of phenological data will allow precise mapping of functional changes in microclimatic variation. The CI will include front end, repository, and back end visualization and analysis tools to facilitate these wider sources of phenology data along with larger agricultural and ecological field studies. All data will be instantly available on public servers (either at iPlant or after pass through) for users to incorporate into their research in scientific studies. This will allow mapping of adaptation and migration of plant communities in response to climate change to serve both this proposal and efforts of the BIEN group.

We will disseminate the CI education and outreach tools developed through these projects into citizen science programs associated with Society for Amateur Scientists (<http://www.sas.org/>) such as IDoScience.net and Labrats (see collaborative email from Dr. Sheldon Greaves). IDoScience.net combines social networking with a powerful database to let citizen scientists collect observations on all aspects of the natural world. LabRats brings together students in grades 6 through 12 once each week to take part in hands-on science-based activities, and take monthly field trips. These programs have already developed the software for users to upload images with an image filter for quality and image size along with the ability to perform statistical analyses on data, but these projects are limited in CI resources. The CI tools developed by our project can be used by citizen scientists to identify and predict phenological events or report stress responses in plants.

Finally, many of the needed modeling calculations are highly intensive, numerically. As noted previously, the NSF-funded Berkeley Open Infrastructure for Network Computing (BOINC) permits such calculations to be distributed over the web to potentially large numbers of computers whose time is “donated” when not in use. A market study commissioned by Welch from the KSU College of Business has shown that an attractive website with interesting, and informative materials is key to recruiting such donations. This would integrate very well with web delivery of the materials described above.

Because BOINC methods require the (in this case) models of interest to reside on contributor machines, it will be easy to add simple interfaces to allow those models to be used in standalone mode for educational purposes. This offers a “closed-loop” scenario where, for instance, (i) citizen scientists collect information on phenology and stress which (ii) flows into the CI for research use, but (iii) also permits 6-12 grade students to locally compare and verify the outputs of models that they run in school as educational modules, (iv) said models automatically reverting to research use in a BOINC calculation when students leave for home where they (v) join their LabRat group and collect/analyze more data.

Describe the GC Team (i) who would interact regularly with iPlant personnel, (ii) projected personnel needs for iPlant outsourced deliverables, and for (iii) evaluation/usability assurance.

The affiliations, research, and outreach particulars of each Principal team member are described on the Cover Page. Here we list individual responsibilities to be outsourced by iPlant, cross reference RFP-mandated activities and/or computational deliverables described earlier, and estimate staffing needs.

Justin Borevitz: Field-to-database pipeline of geo-referenced, phenological and environmental observations suitable for research and/or citizen science. (Activity E & A, as relates. Staffing: GRA.)

Edward Buckler: Integrating (to the level of virtual fusion) statistical, network, and ecophysiological models for project-wide applications. (Activity G: Staffing: project-wide postdoc.)

Melanie Correll: Developing an ecophysiological model of *A. thaliana* and coordinating EOT efforts in concert with project participants and iPlant staff. (Activity G. Staffing: GRA).

Chad Myers: Developing network inference tools, esp. incl. interfacing/exploitation of CI resources and models. (Activity D. Staffing: GRA.)

Johanna Schmitt: RFP-mandated project-wide CI usability assurance in applying integrated phenology models and data to climate change and plant adaptation studies. (Staffing: project-wide postdoc)

Stephen Welch: Contributing to *A. thaliana* ecophysiological modeling, model fusion, text mining, and climate applications. (Activities C & G. Staffing for project wide project evaluation activity, see below.)

Jeffrey W. White: Integrating ecophysiological models and genomic information, coordinating collation of existing crop-related observations and databases. (Activities A, E, & G; staffing, Technician, 0.5 FTE.)

Project management structure: (i) GC Team and iPC responsibilities for project deliverables (ii) project stages with timeline, (iii) critical decision milestones. Estimated completion date: 12/31/11

The management structure will mirror that of the Ecological Genomics Institute (EGI) of Kansas State University (Welch is a Steering Committee member). The project will consist of workgroups each comprised of plant biologists, computer scientists, and iPlant staff. Project Principals plus iPlant managers will be a Steering Workgroup with 1-hr teleconferences every 14 days and two on-site meetings per year. As in the EGI, a member of the Steering Workgroup will meet with each workgroup, but will not chair it. This promotes vertical information flow but avoids an excessively top-down management culture. Once workgroups are formed, a face-to-face meeting will be organised with iPlant staff and members of other GC Teams with common CI interests. The meeting will coordinate, to the extent compatible with individual GC's, the databases, tools, and choice of middle-ware to use for initial prototyping.

After this, workgroups will meet electronically at least once per month (some more frequently) using the CI groupware described above. Work groups will serve as two-way channels of information, representing plant species, selected user groups, education and outreach audiences, and evaluation teams. Individuals listed in Appendix Table A-3 have agreed to serve in the contexts indicated. All individuals listed have seen nearly complete drafts of this proposal. The number expressing willingness to help is an indication of the degree of community buy-in to the specifics presented herein. While all members will undoubtedly make positive contributions, large response guarantees the existence of a subset whose disproportionate efforts will bring success. Even so, as noted previously, it is critical to sustaining interest and to ultimate CI acceptance that the user community perceive value in the system from the beginning. So, on the project *output* side, work groups will be the primary targets/evaluators of educational material on system operation and many members will be its initial users. On the *input* side, groups will provide feedback on system features, performance, identify/share relevant data sets of data, help/advise in preparing educational materials, participate in project evaluation, etc. The timeline in Table A-2 shows specific deliverables and the respective responsibilities of iPlant staff and the GC Team.

Project progress monitoring and evaluation plan.

Progress monitoring will be the on-going responsibility of the project Principals in concert with appropriate iPlant managers. The project will establish an external evaluation team comprised of selected plant biology community members involved in testing (but not in creating) the CI, representative educational and outreach audiences, iPlant project evaluators, and staff from KSU's Office of Educational Innovation and Evaluation (OEIE). Because of her special responsibility for usability assurance, Schmitt (or designate) will be an *ex officio* observer. Formative and summative evaluations will be grounded in the Kellogg Logic Model (78). Evaluation leadership will reside jointly with iPlant evaluators and OEIE.

The latter has 30 evaluators and staff who have been evaluating university, state, and national projects for nearly 10 years. OEIE adheres to the standards of the American Evaluation Association and the program evaluation standards established by the Joint Committee on Standards for Educational Evaluation.

Potential Intellectual Property constraints (if any are known).

Near exclusive preference will be given to non-proprietary, open-sourced software in the selection of extant tools. Exceptions may arise due to market dominance (e.g., ESRI in GIS software), but even then, options may exist (e.g., ESRI's ArcGIS Explorer). Attention will be paid to the balancing of open-sourcing and data exchange with recognition and pre-publication rights to project-originated materials.

Appendix: Tables

Table A-1: List of example relevant databases. Tabulated jointly with the Stress proposal.

Synergistic proposed iPlant Cyberinfrastructures	
Botanical Info. & Ecol. Network	Will supply the most significant existing sets of vegetation data across North and South America, incorporating data resources for plant plot information and taxonomies, and will encompass millions of species occurrence records.
Photosynthesis	Will supply multi-species molecular, phenotypic, and biophysical datasets
Tree-of-Life	Will supply all information relevant to the phylogeny of green plants and be a critical resource in comparisons between genes, genomes, and species
Databases	
TAIR	www.arabidopsis.org Collaborator on this CI Includes the Arabidopsis sequence and maps, data on gene structure and products, metabolic pathways, gene expression (from Geneinvestigator.), phenology of insertion mutants (in some cases), DNA seed stocks, and genetic and physical markers. Text mining initiatives now under review.
GrainGenes	http://wheat.pw.usda.gov/GG2 A collection of curated datasets on genetic and physical maps, genes, QTL, germplasm and gene expression from Triticeae
SoyBase	www.soybase.org Soybean genetic maps, analysis tools, and links to other information
PlantGDB	www.plantgdb.org Collaborator on this CI Plant species-specific EST and GSS databases with inter-species query capabilities, as well as genome browsing and annotation capabilities.
MaizeGDB	www.maizegdb.org Collaborator on this CI Community database for biological information about the crop plant <i>Zea mays</i> ssp. <i>mays</i> . Genetic, genomic, sequence, gene product, functional characterization and literature references,
Tomato Functional Genomics Database	http://ted.bti.cornell.edu/ Expression data, metabolic data, and small RNAs
SOL Genomics Network	http://www.sgn.cornell.edu/ Collaborator on this CI Solanaceae maps, loci, libraries, ESTs, metabolic annotation and phenotypes accessions
BioGRID	http://www.thebiogrid.org Contains protein and genetic interactions from major model organisms.
Phytozome	www.phytozome.org In addition to Gramene database described below, Phytozome is the source repository of many sequenced and annotated genomes from plants e.g. Sorghum, Brachypodium, Poplar and lower plants.
Gramene	www.gramene.org Collaborator on this CI QTLs, genetic maps, genetic markers, Pathways (rice, sorghum, maize and other cereals), genetic diversity and genotype-phenotype associations, Genomes, whole genome alignments, gene-orthologs, ontologies and Comparative maps. All plants served with emphasis on rice, wild rice species, maize, sorghum, poplar and grape.
GEO Express	http://www.ncbi.nlm.nih.gov/geo/ A gene expression/molecular abundance repository supporting MIAME compliant data submissions, and a curated, online resource for gene expression data browsing, query and retrieval

Plants Database	http://plants.usda.gov/ Includes names, plant symbols, checklists, distributional data, species abstracts, characteristics, images, crop information, automated tools, onward Web links, and references.
Massive Arabidopsis Genotype/Phenotype Dataset	http://plants.usda.gov/ 1,000 accessions of <i>A. thaliana</i> genotyped for 250,000 SNPs; 6,000 accessions genotyped for 149 SNPs; 100 phenotypes for 192 genotyped accessions. Collaborator on this CI
Coupled Model Inter-comparison Project (phase 3)	http://www.pcmdi.llnl.gov/ipcc/about_ipcc.php Global climate model output database used for determining climate change science and impacts in the IPCC 4 th Assessment Report
North American Regional Climate Change Assessment Program	http://www.narccap.ucar.edu/ Regional climate model output database for determining climate change science and impacts at regional scales for North America for the IPCC 5 th Assessment Report
The Soil Data Mart	http://soildatamart.nrcs.usda.gov The central repository of soil data operated by the USDA Natural Resources Conservation Service.
TIGR Solanaceae Genomics Resource	http://www.tigr.org/tigr-scripts/tdb/sol/study/sol_study.pl Gene expression data (finished in Sept 2007 now and Archive)
Public Phenology Database	http://www.sws-wis.com/lifecycles/dbase.html View/enter phenological observations from nature.
1001 Arabidopsis Genomes	http://1001genomes.org The complete sequence data for 1001 inbred lines. The data will become available as the project progresses and will be complete within two years. Collaborator on this CI
Chemical Entities of Biological Interest (<i>ChEBI</i>)	http://www.ebi.ac.uk/chebi/ ChEBI encompasses an ontological classification, whereby the relationships between molecular entities or classes of entities and their parents and/or children are specified.
MetaCyc	www.metacyc.org A resource for peer reviewed and published reference metabolic pathways and compounds for all organism Collaborator on this CI
GRIN National Plant Germplasm System of USA	http://www.ars-grin.gov/npgs/ It is the largest repository of germplasm information in the USA
Maize Diversity project	www.panzea.org Germplasm, genotype and phenotype associations from maize
Sorghum Diversity project	http://sorghumdiversity.org/sorghum/index.html Germplasm, genotype and phenotype associations from sorghum
Plant metabolic Network (PMN)	www.plantcyc.org A resource for peer reviewed and published reference metabolic pathways and compounds for plants. Collaborator on this CI
Rice Diversity	www.oryzasnp.org/ OryzaSNP project Collection of SNPs from diverse rice cultivars
Rice Diversity project	http://www.ricediversity.org/ A collaborative effort to explore the genetic basis of variation in rice and its wild ancestors
Gramene's Rice diversity datasets	http://www.gramene.org/db/diversity/diversity_view A repository of rice genetic diversity datasets.
State Ag Exp Station Crop Trials	In electronic & paper form: Will require collation
Project Budburst	http://www.windows.ucar.edu/citizen_science/budburst/ Citizen science for observations of phenological events such as the first bud burst, first leafing, first flower, and seed or fruit dispersal of a diversity of tree, shrub and flower species, including weeds and ornamentals.
National Sustainable Agriculture Research	http://www.attra.org/attra-pub/phenology.html Bloom times, floral calendars, information on phenology

US National Phenology Network	http://www.usanpn.org/ Continental-scale network focused on phenological observations of a few regionally appropriate native species and nationally-cultivated indicator plants	Collaborator on this CI
Natures Calendar (UK)	http://www.naturescalendar.org.uk/ UK historic data and online data entry will both be available.	
GENEVESTIGATOR	https://www.genevestigator.ethz.ch Mine data from Affymetrix Genechips for specific genes or specific conditions.	
Plant Expression Database (PlexDB)	www.plexdb.org PLEXdb (P lant E xpression D atabase) is a unified public resource for gene expression for plants and plant pathogens. Allows comparative analysis of of large-scale expression profiling data sets.	
Nottingham Arabidopsis Stock Centre Transcriptomics Service	http://affymetrix.arabidopsis.info/narrays/experimentbrowse.pl Depository for Affymetrix microarray experiments	
IRIS / IMIS / IWIS	http://seeds.irri.org/iris/ http://seeds.irri.org/imis/ http://seeds.irri.org/iwis/ Respectively, rice, maize, and wheat implementations of the International Crop Information system containing germplasm pedigree, phenotype, and genotype information.	Collaborator on this CI
Ontologies	http://www.geneontology.org/ An ontology of molecular functions, biological processes and cellular locations	
	www.plantontology.org An ontology of plant part anatomy and morphology, along with growth and developmental stages	
	http://www.gramene.org/plant_ontology/ontology_browse.html#to Entry point to a collection of ontologies on traits, genes, plants, environments, taxonomy, and (prototype) geographic place names.	
	http://bioontology.org/wiki/index.php/PATO:Main_Page Phenotype and Attribute Ontology_(PATO) is an ontology of phenotypic qualities, intended for use in a number of applications, primarily defining composite phenotypes and phenotype annotation.	
Virtual Plant	www.virtualplant.org Provides an integrated view of genomic data, expression and gene networks.	Collaborator on this CI
ArrayExpress at the European Bioinformatics Institute	http://www.ebi.ac.uk/microarray-as/ae/ Transcriptomics data, with MIAME- and MINSEQE-compliant data following MGED recommendations. The ArrayExpress Warehouse stores gene-indexed expression profiles from a curated subset of experiments in the repository.	

Table A-2: Project timeline. Lighter and darker green show the approximate division of labor between the GC Team and iPlant, respectively. Purple activities relate to the User GUI. (Users may use off-line visualization tools during 2009 to enable rapid prototyping.) The remaining activities are data-related. Light yellow is for networks and modeling. Pinkish orange denotes CI components (“services”) that support end user “tools”. Text mining tools (pastel green) extract new data and store them and may rely on inferencing and statistical services when processing literature.

Time line	2009				2010				2011			
	A	M	JJA	SOND	JFM	AMJ	JAS	OND	JFM	AMJ	JAS	OND
Preparation stage												
Establish work groups	[Timeline bar: Light green from A to M 2009]											
Finalize tools, databases, CI features	[Timeline bar: Light green from A to M 2009]											
Select prototyping middle-ware	[Timeline bar: Light green from A to M 2009]											
Middle-ware services enhancement stage												
Ontology selection and integration (A: ST)	[Timeline bar: Light green from A to M 2009, Darker green from M 2009 to JAS 2010]											
Resource allocation and execution mgt services	[Timeline bar: Darker green from JJA 2009 to JAS 2010]											
Data access services	[Timeline bar: Darker green from M 2009 to JAS 2010]											
Tools and algorithms access services	[Timeline bar: Darker green from JJA 2009 to JAS 2010]											
Results presentation services	[Timeline bar: Darker green from JFM 2010 to JAS 2010]											
End-user services / tools implementation stage												
User communication services (H: Immediate)	[Timeline bar: Light green from A to M 2009, Darker green from M 2009 to JAS 2010]											
Visualization services	[Timeline bar: Light green from M 2009 to JAS 2010, Darker green from JAS 2010 to JAS 2011]											
"Omics" data fusion services	[Timeline bar: Darker green from JJA 2009 to JAS 2010]											
Data integration tools (B: IT)	[Timeline bar: Darker green from JFM 2010 to JAS 2010]											
Integrate species, landraces, genotype data (E: ST-PST)	[Timeline bar: Light green from M 2009 to JAS 2010, Darker green from JAS 2010 to JAS 2011]											
Establish metabolism data resource (F: ST)	[Timeline bar: Light green from M 2009 to JAS 2010, Darker green from JAS 2010 to JAS 2011]											
Text mining tools (C: ST, IT, LT)	[Timeline bar: Light green from M 2009 to JAS 2010, Darker green from JAS 2010 to JAS 2011]											
Modeling, inference, and statistical services	[Timeline bar: Light green from M 2009 to JAS 2010, Darker green from JAS 2010 to JAS 2011]											
Network inference tools (D: PST)	[Timeline bar: Light green from M 2009 to JAS 2010, Darker green from JAS 2010 to JAS 2011]											
Quantitative modeling (G: ST, IT, LT)	[Timeline bar: Light green from M 2009 to JAS 2010, Darker green from JAS 2010 to JAS 2011]											
Education, outreach, training, and evaluation stage												
EOT materials development (All stages)	[Timeline bar: Light green from M 2009 to JAS 2010, Darker green from JAS 2010 to JAS 2011]											
Usability assurance and evaluation (All stages)	[Timeline bar: Light green from M 2009 to JAS 2010, Darker green from JAS 2010 to JAS 2011]											

Table A-3. Examples of potential phenology working groups and suggested initial members.
Final membership and scope will be defined by their members. Sample confirmations are annexed after the citations.

Last name	First name	Institute	Country	Species	Mod	Algo	Gen	Gen	Ecol	Ecop	CI	Data	Clim	Outr	Interests
GENOMICS AND GENETICS															
Beavis	Bill	Iowa State University	USA	MZ			X	X							Finding the genotypic basis of complex traits for purposes of developing high throughput diagnostic tests.
Beebe	Steve	Centro Internacional de Agricultural Tropical (CIAT)	Colombia	GL				X							Bean breeder working on adaptation to abiotic stresses, biofortification, disease and pest resistance.
Bergelson	Joy	University of Chicago	USA	AT			X	X	X						Development of tools and resources to allow for the mapping of ecologically important genes in <i>A. thaliana</i>
Blair	Matthew	Centro Internacional de Agricultural Tropical (CIAT)	Colombia	GL			X	X							Bean breeder and molecular biologist
Borevitz	Justin	University of Chicago	USA	AT,SI				X	X					X	Quantitative/population genetics are used to dissect phenotypic variation in <i>Arabidopsis</i> and switchgrass
Buckler	Ed	USDA/ARS Plant, Soil, and Nutrition Research Unit	USA	MZ			X	X		X					Basis of complex traits with an emphasis on dissecting abiotic stress tolerance in maize
Castro	Ariel	Oregon State University	USA	BA				X							Breeding, genetics, physiology
Correll	Melanie	University of Florida	USA	AT	X			X						X	Ecophysiological modeling of <i>Arabidopsis</i> ; education, outreach and training
Coupland	George	Max Planck Institute of Plant Breeding	Germany	AT					X						Extensive multi-environment flowering phenotype data; increasing amounts of quantitative RNA & protein data
Dreisigacker	Susanne	CIMMYT	Mexico	WH	X		X	X							CIMMYT wheat molecular biologist
Dubcovsky	Jorge	University of California at Davis	USA	WH			X	X						X	Wheat genomics and breeding with emphasis on phenology
Harmer	Stacey	University of California at Davis	USA	AT				X							Circadian clock studies in <i>Arabidopsis</i>
Hunt	L.A. (Tom)	University of Guelph	Canada	WH, BA	X			X							Wheat breeder and ecophysiologicalist who has developed two wheat models.
Jaradat	Abdullah	USDA-ARS North Central Soil Conservation Research Lab	USA	WH, MZ, SG, GL	X			X						X	Farming systems and climate change
Juenger	Thomas	University of Texas	USA	AT				X	X	X					Physiological and gene expression responses to drying soil in <i>Arabidopsis</i>
Keurentjes	Joost	Wageningen University and Research Centre	Netherlands					X		X		X			Quantitative genetics of complex traits and plant metabolism
Kliebenstein	Daniel	University of California at Davis	USA	AT				X		X					Methods of understanding the quantitative genetic architecture of complex traits; secondary metabolites in <i>Arabidopsis</i>
Lauter	Nick	USDA/ARS Corn Insects and Crop Genetics Research Unit	USA	MZ			X	X				X			QTL dissection of complex traits; juvenility in maize; database tools for functional genomics
Maloof	Julin	University of California at Davis	USA	AT				X	X						Molecular and QTL studies of light-regulated growth
Maroof	Sagai	Virginia Tech	USA	SB			X								Grain legume (esp. soybean) genomics
Nordborg	Magnus	University of Southern California	USA	AT			X	X	X						Combines approaches from population genetics, statistical genetics, and molecular evolution to study adaptation
Payne	Tom	CIMMYT	Mexico	WH				X				X			Curator of CIMMYT's wheat genetic resources

Table A-3. Examples of potential phenology working groups and suggested initial members.
Final membership and scope will be defined by their members. Sample confirmations are annexed after the citations.

Last name	First name	Institute	Country	Species	Mod	Algo	Gen	Gene	Ecol	Ecop	CI	Data	Clim	Outr	Interests
Pires	Chris	University of Missouri	USA				X	X	X						Evolutionary biology, esp. molecular systematics and comparative genomics with a focus on the evolution and ecology of polyploid plants. Agronomy of cabbage, broccoli, cauliflower, canola, onions, asparagus, irises, orchids
Rane	Jagadish	CIAT	Colombia	RI				X							Promoter and gene combinations and their effect on the field behavior of rice under water stress
Rooney	Bill	Texas A&M University	USA	SG			X	X							Sorghum breeding and genetics
Schmitt	Johanna	Brown University	USA	AT	X			X	X				X		Evolutionary studies of Arabidopsis flowering time as related to climate change
von Korff	Maria	Max Planck Institute of Plant Breeding	Germany	BA				X							Studies flowering in barley using mapping and transgenic lines.
Udvardi	Michael	The Samuel Roberts Noble Foundation	USA	GL			X	X	X						Plant/soil macronutrient relations; environmental stress; storage metabolism during seed development; gene regulatory networks governing nodule development and differentiation in <i>Medicago truncatula</i>
Weigel	Detlef	Max Planck Institute for Developmental Biology	Germany	AT			X	X	X						Integrating genetic networks and trait variation to understand adaptive change, and to predict species performance in changing environments.
Weinig	Cynthia	University of Wyoming	USA	AT				X	X					X	Shade avoidance mechanisms in Arabidopsis; K-12 Outreach
Willis	John	Duke University	USA	MG				X	X						Genetic variation and QTL analysis of flowering responses to photoperiod, vernalization
ALGORITHMS															
Caragea	Doina	Kansas State University	USA			X					X	X			Machine learning, heterogenous databases
Das	Sanjoy	Kansas State University	USA			X									Optimization
Liao	Li	University of Delaware	USA		X	X									Development of statistical and machine learning methods that incorporate domain specific biological knowledge
Myers	Chad	Univ. Minnesota, Minneapolis	USA			X	X				X				Integration and inference from heterogeneous genomic data, applying machine learning and data mining to understand gene function and model biological networks
Rawlings	Chris	Rothamsted Research	UK			X					X	X			Head, Dept Biomathematics and Bioinformatics, Rothamsted Research. As a group, researchers in this unit have interests touching on many aspects of the proposed cyberinfrastructure.
Yan	Xifeng	University of California at Santa Barbara	USA			X									Graph mining and management algorithms
CLIMATOLOGY															
Goheer	Arif	Global Change Impact Studies Centre (GCISC), Quaid-e-Azam University	Pakistan										X		
Hijmans	Robert	University of California at Davis	USA	MZ							X		X		Agroclimatologist, biogeographer and developer of widelyused DIVA-GIS
Hoogenboom	Gerrit	University of Georgia	USA	GL, SG, CO	X								X	X	Climatologist, ecophysiologicalist and developer of DSSAT modeling platform
Koo	Jawoo	International Food Policy Research Institute (IFPRI)	USA	MZ, PN, MI, SG									X		Climatology in Sub-Saharan Africa; ecophysiological models
Martin	Nicolas		USA	MZ	X						X		X		

Species codes: AT = Arabidopsis; BA = Barley; CO = Cotton; GL = Grain legumes; MG = Mimulus (monkey flower); MZ = Maize; RI = Rice; SC = Sugarcane; SG = Sorghum; SI = Switchgrass; WH = Wheat Tomato; WB = Wheat and barley.

Table A-3. Examples of potential phenology working groups and suggested initial members.

Final membership and scope will be defined by their members. Sample confirmations are annexed after the citations.

Last name	First name	Institute	Country	Species	Mod	Algo	Gen	Gene	Ecol	Ecop	CI	Data	Clim	Outr	Interests
Niyogi	Dev	Purdue University	USA										X		Research focus is on studying and representing land surface processes and effects of vegetation – atmosphere interactions on environmental processes at different scales.
Rolando (Rivera)	Roger	Centro Meteorológico de Camaguey, INSMET	Cuba	RI, SC, MZ	X								X		
Russo	Susana	Instituto de Suelos, INTA	Argentina	WH, MZ, GL	X								X		
Takle	Gene	Iowa State University	USA	AT	X								X		Evaluating effects of climate changes on agroecosystems using climate-crop coupled models
Weiss	Al	University of Nebraska	USA	WH, MZ, GL	X								X		Agroclimatologist; growth and development of winter wheat, including tillering and kernel nitrogen.
Yun	Jin I.	Kyung Hee University	South Korea	RI, WH	X								X		Estimating climate change impacts.
CYBERINFRASTRUCTURE															
Bruskiewich	Richard	IRRI	Philippines	RI			X					X	X		Bioinformatics of rice, wheat and maize
McLaren	Graham	Bioinformatics and Crop Information Systems Generation Challenge Program	Mexico	RI, WH, MZ, GL								X	X	X	Bioinformatics and crop information systems
Sauro	Herbert	University of Washington	USA									X			Control and dynamics of cellular networks. Theory, algorithms, standards (SBML) and software for simulation and analysis.
Taylor	Stephanie	Colby College	USA		X							X			Mathematical modeling of the circadian clock and numerical methods
Yang	Yubin	Texas A&M Agrilife Research & Extension Center	USA	RI, MZ, CO	X							X			Population Dynamics, Systems Analysis, Mathematical Modeling and Computer Simulation of Agroecosystems, Agricultural Decision Systems.
DATABASES															
Huala	Eva	The Arabidopsis Information Resource	USA	AT								X			Director of TAIR
ECOLOGY AND EVOLUTION															
Gepts	Paul	University of California at Davis	USA	GL			X	X	X						Evolution and domestication of grain legumes
McClung	Rob	Dartmouth	USA	AT					X						Circadian clock studies
ECOPHYSIOLOGY															
Boote	Ken	University of Florida	USA	GL, MZ	X							X			Ecophysiology of grain legumes, maize and other crops. Developer of CROPGRO.
Korekar	Girish	Field Research Laboratory, Defence and Research Development Organisation	India									X			Seabuckthorn and Apricot phenology
Krishnan	Prameela	Central Rice Research Institute	India	RI											
McMaster	Greg	USDA-ARS, Agricultural Systems Research Unit	USA	WH, SG, MZ	X							X		X	Crop phenology. Decision support tools for producers.
Mukhtar	Ahmad	University of Arid Agriculture (Rawalpindi)	Pakistan												
Mullet	John	Texas A&M University	USA	SG											Genetic basis of plant drought tolerance, sorghum genomics, and the genetic improvement of bioenergy crops.

Species codes: AT = Arabidopsis; BA = Barley; CO = Cotton; GL = Grain legumes; MG = Mimulus (monkey flower); MZ = Maize; RI = Rice; SC = Sugarcane; SG = Sorghum; SI = Switchgrass; WH = Wheat Tomato; WB = Wheat and barley.

Table A-3. Examples of potential phenology working groups and suggested initial members.
Final membership and scope will be defined by their members. Sample confirmations are annexed after the citations.

Last name	First name	Institute	Country	Species	Modd	Algo	Gen	Gen	Ecol	Ecop	CI	Data	Clim	Outr	Interests
Murari	Kamal	The Energy and Resources Institute	India												
Rao	Idupulapati	Centro Internacional de Agricultural Tropical (CIAT)	Colombia	GL						X					Ecophysiology of plant adaptation to abiotic stresses
Traore	Pierre_C.	ICRISAT	Mali	SG	X					X			X	X	Sorghum phenotyping data, modeling and geospatial expertise focused on the West African subhumid and semiarid conditions.
White	Jeff	USDA ARS, Arid Land Agricultural Res. Center	USA	WH, SG, GL	X					X		X		X	Adaptation of plants to climate change
EDUCATION & OUTREACH															X
Correll	Melanie	University of Florida	USA	AT, GL	X		X							X	Integrating bioinformatics and ecophysiological modeling to predict plant adaptation.
Gallo	Maria	University of Florida	USA	GL			X	X						X	Peanut genomics in applications to breeding
Greaves	Sheldon	Society of Amateur Scientists	USA											X	Outreach to large numbers of active citizen scientists engaged in a broad range of activities
Horn	Lydia N.		Namibia		X									X	
Levey	Doug	University of Florida	USA											X	Chair for the Science Partners in Inquiry-based Collaborative Education (SPICE) program at the UF
Ottman	Mike	University of Arizona	USA	WH, BA, SG	X									X	Wheat, barley and sorghum physiology and agronomy. Outreach to producers and other stakeholders.
Sarkar	Reshmi	Lincoln University	USA	RI, MZ	X									X	Ecophysiological modeling of rice and maize. End-user testing.
MODELS															
Brunel	Sophie	INRA Bordeaux-Aquitaine	France	MZ	X										Simulation of plant response to nutrients with emphasis on P deficiencies.
Guisard	Yann	Charles Sturt University	Australia		X										Three dimensional modelling of biophysical processes in orchards and vineyards, with emphasis on stomatal behaviour, predictive growth and visualisation. Field of application include tertiary education and irrigation strategies.
Jones	Jim	University of Florida	USA	GL, CO, TO	X						X			X	Agricultural and biological systems analysis.
Kemanian	Armen	Texas AgriLife Research, Blackland Research and Extension Center (BREC)	USA	SG, WH, MZ	X									X	Agricultural and biological systems analysis.
Muller	Johannes	Martin-Luther-University Halle-Wittenberg	Germany	BA	X						X				interactions between processes of plant development and formation of dry mass and yield
Porter	John	Royal Agricultural and Veterinary University (KVA)	Denmark	WH	X										Crop ecology and physiology, biological modelling and agro-ecology
Sanchez	Ricardo	Kean University	USA	MZ, TO, BR	X	X								X	Systems Simulation with emphasis on commercial applications. Understanding how theoretical approaches of the models relate to mathematical representations, especially in relation to systems biology.
Shanker	Arun	Central Research Institute for Dryland Agriculture, ICAR	India	WH	X										Physiology
Singels	Abraham	South African Sugarcane Research Institute	South Africa	SC	X									X	Sugarcane ecophysiology and modeling

Species codes: AT = Arabidopsis; BA = Barley; CO = Cotton; GL = Grain legumes; MG = Mimulus (monkey flower); MZ = Maize; RI = Rice; SC = Sugarcane; SG = Sorghum; SI = Switchgrass; WH = Wheat Tomato; WB = Wheat and barley.

Table A-3. Examples of potential phenology working groups and suggested initial members.
Final membership and scope will be defined by their members. Sample confirmations are annexed after the citations.

Last name	First name	Institute	Country	Species	Mod	Algo	Gen	Gen	Ecol	Ecop	CI	Data	Clim	Outr	Interests
Swain	Dillip		India	RI	X					X				X	Modeling of climate change impacts on rice production
Thorp	Kelly		USA	WH, CO	X					X					Interactions of phenology and crop water use. Precision agriculture.
Wang	Sam	University of Arizona (Maricopa Agricultural Center)	USA	WH, BA, SG, GL	X									X	Agronomy and modeling of crop-weed interactions
Welch	Steve	Kansas State University	USA	AT	X	x					X				Mathematical models of plant growth and development
Wilson	L.T. (Ted)	Texas A&M Agrilife Research & Extension Center	USA	RI, CO	X						X				Croppiing systems ecology and modeling
Yin	Xinyou	Wageningen Agricultural University	Netherlands	RI, BA	X						X				Linking ecophysiological modeling with functional genomics, in relation to climate variability (especially to drought stress).
Zhu	Xinguang	Partner Institute of Computational Biology, CAS	China		X										Mechanistic models of photosynthesis

Citations

1. Morin, X., C. Augspurger and I. Chuine, 2007 Process-based modeling of species' distributions: What limits temperate tree species' range boundaries? *Ecology* 88: 2280-2291.
2. Ward JK, Antonovics J, Thomas RB, Strain BR. 2000. Is atmospheric CO₂ a selective agent on model C₃ annuals. *Oecologia* 123:330-341.
3. Foley JA, Prentice C, Ramankutty N, Levis S, Pollard D, Stith S, Haxeltine A. 1996. An integrated biosphere model of land surfaces processes, terrestrial carbon balance, and vegetative dynamics. *Global Biogeochemical Cycles* 10:603-628.
4. Yan W, Wallace DH. 1998. Simulation and prediction of plant phenology for five crops based on photoperiod x temperature interaction. *Annals of Botany* 81: 705-716.
5. Grimm SS, Jones JW, Boote KJ, Hesketh JD. 1993. Parameter estimation for predicting flowering date of soybean cultivars. *Crop Science* 33: 137-144.
6. Piper EL, Boote KJ, Jones JW, Grimm SS. 1996. Comparison of two phenology models for predicting flowering and maturity date of soybean. *Crop Science* 36: 1606-1614.
7. Hadley P, Roberts EH, Summerfield RH. 1983. A quantitative model of reproductive development in cowpea [*Vigna unguiculata* (L.) Walp.] in relation to photoperiod and temperature and implications for screening germplasm. *Annals of Botany* 51: 531-43.
8. Summerfield RJ, Ellis RH, Craufurd PQ. 1996. Phenological adaptation to cropping environment. From evaluation descriptors of times to flowering to the genetic characterization of flowering responses to photoperiod and temperature. *Euphytica* 92: 281-86.
9. Wallace DH, Enriquez GA. 1980. Day length and temperature effects on days to flowering of early and late maturing beans (*Phaseolus vulgaris* L.). *Journal American Society For Horticultural Science* 105: 583-91.
10. Welch SM, Roe JL, Das S, Dong Z, He R, Kirkham MB. 2005. Merging genomic control networks and soil-plant-atmosphere-continuum (SPAC) models. *Agricultural Systems* 86: p. 243-274.
11. Yan W, Hunt LA. 1999a. An equation for modeling the temperature response of plants using only the cardinal temperatures. *Annals of Botany* 84: 607-614.
12. Yan W, Hunt LA. 1999b. Reanalysis of vernalization data of wheat and carrot. *Annals of Botany* 84: 615-619
13. Wilczek AM, Roe JL, Knapp MC, Cooper MD, Lopez- Gallego C, Martin LJ, Muir CD, Sim S, Walker A, Anderson J, Egan JF, Moyers B, Petipas R, Giakountis A, Charbit E, Coupland G, Welch SM, Schmitt J. 2009. Effects of genetic perturbation on seasonal life history plasticity. *Science*. (Published online in *Science Express* 15 Jan 09).
14. Boss PK, Bastow RM, Mylne JS, Dean C. 2004. Multiple pathways in the decision to flower: enabling, promoting, and resetting. *Plant Cell* 16 Suppl: S18-31.
15. Henderson IR, Dean C. 2004. Control of *Arabidopsis* flowering: the chill before the bloom. *Development* 131: 3829-38

16. Putterill J, Laurie R, Macknight R. 2004. It's time to flower: the genetic control of flowering time. *Bioessays* 26: 363-373.
17. Simpson GG. 2004. The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of *Arabidopsis* flowering time. *Current Opinion in Plant Biology* 7: 570-4.
18. Bernier G, Perilleux C. 2005. A physiological overview of the genetics of flowering time control. *Plant Biotechnology Journal* 3: 3-16.
19. He Y, Amasino RM. 2005. Role of chromatin modification in flowering-time control. *Trends in Plant Science* 10: 30-5.
20. Vandenbussche, F., R. Pierik, F.F. Millenaar, L.A.C.J. Voesenek, and D. Van Der Straeten, Reaching out of the shade. *Current Opinion in Plant Biology*, 2005. 8: p. 462-468.
21. Kramer BP, Fussenegger M. 2005. Hysteresis in a synthetic mammalian gene network. *PNAS* 102:9517-9722.
22. Brandman O, Ferrell JE, Li R, Meyer T. 2005. Interlinked fast and slow positive feedback loops drive reliable cell decisions. *Science* 310:496-498.
23. Elowitz MB, Leibler S. 2000. A synthetic oscillatory network of transcriptional regulators *Nature* 403:335-338.
24. Jack T. 2004. Molecular and genetic mechanisms of floral control. *Plant Cell* 16, S1-17.
25. Hempel FD, Welch DR, Feldman LJ. 2000. Floral induction and determination: where is flowering controlled? *Trends in Plant Science* 5, 17-21.
26. Sablowski R. 2007. *Journal of Experimental Botany* 58, 899.
27. Ratcliffe OJ, Bradley DJ, Coen ES. 1999. Separation of shoot and floral identity in *Arabidopsis*. *Development* 126: 1109-1120.
28. Blazquez MA. 1997. Illuminating flowers: CONSTANS induces LEAFY expression. *Bioessays* 19, 277-279.
29. Davidson, E., et al. 2002. A genomic regulatory network for development. *Science*, 2002. 295:1670-1678.
30. Salisbury F. 1963. *The flowering process*. Pergamon Press, Oxford, New York.
31. Thomas B, Vince-Prue D. 1997. *Photoperiodism in plants*. Academic Press, San Diego.
32. Springer CJ, Orozco RA, Kelly Jk, Ward JK. 2008. Elevated CO₂ influences the expression of floral-initiation genes in *Arabidopsis thaliana*. *New Phytologist* 178:63-67.
33. Wang JY. 1960. A critique of the heat unit approach to plant response studies. *Ecology* 41:785-90.
34. White, JW, Hoogenboom G. 1996. Simulating effects of genes for physiological traits in a process-oriented crop model. *Agron. J.* 88:416-422.

35. Messina, CD, Jones JW, Boote KJ, Vallejos CE. 2006. A gene-based model to simulate soybean development and yield responses to environment. *Crop Sci.* 46:456-466.
36. White, JW, et al. 2008. Simulation-based analysis of effects of Vrn and Ppd loci on flowering in wheat. *Crop Sci* 48:678-687.
37. Zeilinger MN, Farre EM, Taylor SR, Kay SA, Doyle FJ. 3rd. 2006. A novel computational model of the circadian clock in Arabidopsis that incorporates PRR7 and PRR9. *Mol Syst Biol* 2:58.
38. Locke, J.C. et al. 2006. Experimental validation of a predicted feedback loop in the multi-oscillator clock of Arabidopsis thaliana. *Mol Syst Biol* 2:59.
39. Welch SM, Roe JL, Dong Z. 2003. A genetic neural network model of flowering time control in Arabidopsis thaliana. *Agronomy Jour.* 95:71-81.
40. IPCC, 2007: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller (eds)] Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp
41. Cooper, M., D.W. Podlich, and O.S. Smith. 2005. Gene to phenotype models and complex trait genetics. *Australian Journal of Agricultural Research.* 56:895-918.
42. NRC. 2008. Achievements of the National Plant Genome Initiative and new horizons in plant biology. www.nap.edu/catalog.php?record_id=12054.
43. Zhu XG, de Sturler E, Long SP. 2007. Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: a numerical simulation using
44. an evolutionary algorithm. *Plant Physiol* 145:513-26.
45. Gutierrez, R.A. et al. 2008. Systems approach identifies an organic nitrogen-responsive gene network that is regulated by the master clock control gene *CCA1*. *PNAS* 105:4939-4944.
46. Hammer, G. et al. 2006. Models for navigating biological complexity in breeding improved crop plants. *Trends Plant Sci* 11:587-93.
47. Espinosa-Soto C, Padilla-Longoria P, Alvarez-Buylla ER. 2004. A gene regulatory network model for cell-fate determination during Arabidopsis thaliana flower development that is robust and recovers experimental gene expression profiles. *Plant Cell* 16:2923-2939.
48. Smith R.S. et al. 2006. A plausible model of phyllotaxis. *PNAS* 103:1301-1306.
49. Nakagawa, H., J. Yamagishi, N. Miyamoto, M. Motoyama, M. Yano, and K. Nemoto. 2005. Flowering response of rice to photoperiod and temperature: a QTL analysis using a phenological model. *Theor Appl Genet*, 110: 778-86.
50. Reymond, M., et al. 2003. Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. *Plant Physiol.* 131:664-675.
51. Tardieu. 2003. Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. *Trends in Plant Science* 8:9-14.

52. Tardieu F, et al. 2005. Linking physiological and genetic analyses of the control of leaf growth under fluctuating environmental conditions. *Australian Journal of Agricultural Research*. 56:937-946.
53. Hanks J, Ritchie J. 1991. *Modeling plant and soil systems*. ASA, CSSA, SSSA, Madison, WI.
54. Jones, J.W., G. Hoogenboom, C.H. Porter, K.J. Boote, W.D. Batchelor, L.A. Hunt, P.W. Wilkens, U. Singh, A.J. Gijssman, and J.T. Ritchie. 2003. The DSSAT Cropping System Model. *European Journal of Agronomy* 18:235-265.
55. Cline et al. 2007. Integration of biological networks and gene expression data using Cytoscape. *Nat. Protocols* 2(10):2366-2382.
56. Wei X, Croft B, McCallum A. 2006. Table extraction for answer retrieval. *Information Retrieval Journal* 9(5):589-611.
57. Pinto D, McCallum A, Wei X, Croft B. 2003. Table Extraction Using Conditional Random Fields. *Proceedings of the ACM SIGIR*.
58. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. 2003. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* 31(1):258-261.
59. Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, Emili A, Snyder M, Greenblatt JF, Gerstein M. 2003. A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science* 302(5644):449-453.
60. Lee I, Date SV, Adai AT, Marcotte EM. 2004. A probabilistic functional network of yeast genes. *Science* 306(5701):1555-1558.
61. Jaimovich A, Elidan G, Margalit H, Friedman N. 2005. Towards an integrated protein-protein interaction network. *Research in Computational Molecular Biology, Proceedings* 3500:14-30.
62. Myers CL, Robson D, Wible A, Hibbs MA, Chiriac C, Theesfeld CL, Dolinski K, Troyanskaya OG. 2005. Discovery of biological networks from diverse functional genomic data. *Genome Biol*, 6(13):R114.
63. Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q: GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol* 2008, 9 Suppl 1:S4
64. Friedman N. 2004. Inferring cellular networks using probabilistic graphical models. *Science*, 303(5659):799-805.
65. Perrin BE, Ralaivola L, Mazurie A, Bottani S, Mallet J, d'Alche-Buc F. 2003. Gene networks inference using dynamic Bayesian networks. *Bioinformatics*, 19 Suppl 2:ii138-148.
66. Zou M, Conzen SD. 2005. A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics*, 21(1):71-79.
67. Albert I, Thakar J, Li S, Zhang R, Albert R: Boolean network simulations for life scientists. *Source Code Biol Med* 2008, 3:16.

68. Kelley BP, Yuan B, Lewitter F, Sharan R, Stockwell BR, Ideker T. 2004. PathBLAST: a tool for alignment of protein interaction networks. *Nucleic Acids Res*, 32(Web Server issue):W83-88.
69. Flannick J, Novak A, Srinivasan BS, McAdams HH, Batzoglou S. 2006. Graemlin: general and robust alignment of multiple large interaction networks. *Genome Res*, 16(9):1169-1181.
70. Stein, LD. 2008 Wiki features and commenting - Towards a cyberinfrastructure for the biological sciences: progress, visions and challenges. *Nat Rev Genet* 9:678-688.
71. Cannataro M, Talia D. 2003. The knowledge grid. *Communications of the ACM* 46(1):89-93.
72. Mastroianni C, Talia D, Trunfio P. 2004. Metadata for Managing Grid Resources in Data Mining Applications *J. Grid Computing* 2(1):85-102.
73. Congiusta A, Talia D, Trunfio P. 2007. Distributed data mining services leveraging WSRF. *Future Generation Computer Systems* 23(1):34-41.
74. Bulatewicz T, Andresen D, Welch SM, Jin W, Das S, Miller M. 2007. A software system for scalable parameter estimation on clusters. In: *Proc. of 8th LCI International Conference on High-Performance Clustered Computing*.
75. Hammer G, Cooper M, Tardieu F, Welch S, Walsh B, van Eeuwijk F, Chapman S, Podlich D. 2006. Models for navigating biological complexity in breeding improved crop plants. *Trends in Plant Science* 11:587-593.
76. Karwowski & Prusinkiewicz. 2003. *Electronic Notes in Theoretical Computer Science* 86(2):134-152
77. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23(19):2633-2635.
78. WK Kellogg Foundation. 2004. *Logic model development guide*. Battle Creek, Michigan.



Rothamsted Research
Harpenden, Herts, AL5 2JQ

Telephone: (01582) 763133

International: +44 1582 763133

Fax: +44 (0)1582 760981

Web: <http://www.rothamsted.bbsrc.ac.uk/>

Director: Professor I.R. Crute

ROTHAMSTED
RESEARCH

30th January 2009

Prof. Stephen M Welch
Department of Agronomy
Kansas State University
Manhattan, KS, USA 66506

Prof. Ruth Grene
Dept of Plant Pathology, Physiology and Weed Science,
435 Old Glade Road
Virginia Tech
Blacksburg
VA 24061-0330

Centre for Mathematical and Computational Biology

Scientific Director: Professor Chris Rawlings

e-mail: chris.rawlings@bbsrc.ac.uk

Re: *Cyberinfrastructure projects in Support of: “Genetic and Ecophysiological Studies of Plant Phenological Control in Complex and Changing Environments” and “Model-Building In Plant Stress Biology: Effects of Climate Change Factors on Crops”*

Dear Professors Welch and Grene

I wish to express our enthusiasm for the proposals that you are making to the iPlant Collaborative Board. Our research in the Centre for Mathematical and Computational Biology at Rothamsted Research is very closely aligned with the computational biology objectives in both these proposals and I very much hope that we can collaborate with this exciting initiative to create a plant cyberinfrastructure.

I realise that as UK scientists we are not eligible for any direct support from the CI projects, but I believe that it is very important that the project has an international perspective. In the plant and crop science research community around the world we are faced with the same problems after all. At Rothamsted we have both software tools and relevant experience that we would like to contribute to your project and we have contacts and collaborators in computer and plant science that may also be interested but are not yet informed about iPlant. I would therefore like to offer to coordinate a bid for collaborative funding from our research agency (BBSRC) that would enable my colleagues at Rothamsted and our collaborators to become active participants in iPlant projects by supporting travel and subsistence funds. We would also be happy to host iPlant visitors here at Rothamsted for extended periods of time if this were helpful.

I would like to elaborate on where our research has particular relevance to the work you have proposed.

Objectives A, B and C

I would like to draw your attention to work we are conducting on semantic data integration and an open source software system we have developed called Ondex (www.ondex.org). This software uses a graph based approach to integrate and visualise (very) large biological networks of interacting genes, proteins, biochemical reactions and other life science data resources such as ontologies. In a collaboration with the myGrid and Taverna (<http://www.mygrid.org.uk/tools/taverna/>) projects in Manchester we are extending Ondex with workflow and Grid technologies to address the issue of integrating bioinformatics methods. In collaboration with our National Centre for Text Mining (<http://www.nactem.ac.uk/>) we are also introducing more comprehensive methods for text mining. One of the key application cases for Rothamsted under the Ondex project is to use our data integration framework to develop methods for prioritising candidate genes under crop QTLs based on integrated genomic and literature data. We are also very interested in developing plant trait ontologies to facilitate the comparison of QTL intervals.

Objectives D & G

We are developing graph-theoretic methods and adapting approaches from computer science for analysing the connectedness of biological networks. We are also collaborating with the University of Newcastle with a view to integrating statistical methods for inferring biological networks from our data integration platform.

Objective E

With colleagues in our Plant Science departments we are involved in development of methods for curating and integrating crop and plant phenotype and genotype data. For example much of the work developing the databases that underpin one of the key Brassica species genome projects <http://www.brassica.info/resource/databases/cropstore.php> is now undertaken at Rothamsted (with Graham King). We are also making these methods more generally applicable to other crops and in particular to wheat and second generation bioenergy crops (e.g. willow). We also maintain the Multinational Brassica Genome Project gene registry and have ongoing projects that will extend the scope of data we will be managing to include that from eQTL studies.

Objective G

Rothamsted has an established record in the development of quantitative agroecological models including models of crop productivity such as the Sirius wheat model (<http://www.rothamsted.bbsrc.ac.uk/mas-models/sirius.php>) and the Rothamsted soil carbon model RothC (<http://www.rothamsted.bbsrc.ac.uk/aen/carbon/rothc.htm>). These models and others developed here are being used with our stochastic weather simulator <http://www.rothamsted.bbsrc.ac.uk/mas-models/larswg.php> that enables us to predict the likely impacts of future climate change scenarios on agricultural systems.

I hope this information provides you with convincing evidence that Rothamsted Research could make a valuable contribution to an international collaboration on a plant cyberinfrastructure. I very much look forward to working with you in the future and wish you good luck with your proposal.

Yours sincerely,



Professor Chris Rawlings

Rothamsted Research
<http://www.rothamsted.bbsrc.ac.uk/bab/>



UNIVERSITY OF FLORIDA

College of Liberal Arts and Sciences
Department of Biology

PO Box 118525
Gainesville, Florida 32611
Tele: (352) 392-9269
Fax: (352) 392-3704
DLEVEY@zoo.ufl.edu

2 February, 2009

Melanie J. Correll, PhD
Agricultural and Biological Engineering Dept.
University of Florida
209 Frazier Rogers Hall
Gainesville, FL 32611-0570

Dear Dr. Correll:

As the chair for the Science Partners in Inquiry-based Collaborative Education (SPICE) program at the University of Florida, I am pleased to offer collaboration on the education and outreach activities for your proposals "Building Predictive Models for Effects of Pan- and Species-Specific Stress-Related Climate Change Factors on Crop Species" (Greene *et al.*) and "Cyberinfrastructural Support for the Genetic and Ecophysiological Decipherment of Plant Phenological Control in Complex and Changing Environments" (Welch *et al.*) for the iPlant Collaborative. Your project topic on the effects of global climate change on plant flowering and stress responses align directly with the SPICE theme of Ecosystem Health and Sustainability.

SPICE is an NSF GK-12 program, now in its sixth year. Like all GK-12 programs, our emphasis is on training graduate students in broader impacts and in getting youth interested into the STEM (science, technology, engineering, and math) career pipeline. Specifically, we have provided extensive training in inquiry-based teaching techniques to approximately 50 University of Florida graduate students, thus far. These "SPICE Fellows" are paired with experienced middle school science teachers in our county's six most disadvantaged schools (both urban and rural). They spend two days per week for the entire year, guiding students on an exploratory quest about Ecosystem Health. Importantly, they become mentors and role models for the students - they are viewed by the students as scientists, not teachers. (In fact, they wear flashy shirts embroidered with "I AM A SCIENTIST!") Although we have the computer hardware in the classroom we lack the cyberinfrastructure to teach quantitative biology that your project can provide, particularly in the area of global climate change and its effect on plant responses.

Our SPICE graduate student fellows, as well as our Program Coordinator, Suzan Smith who is a middle school teacher for almost 15 years, can provide guidance in the development and give feedback for your educational and outreach tools. We would also be pleased to provide a venue to disseminate and test your tools in middle school

classrooms. The developed lesson plans and learning modules can be made available at the SPICE website for downloading (<http://www.spice.centers.ufl.edu/index.html>).

The collaborative efforts described will allow us to train the SPICE fellows in utilizing cyberinfrastructure to teach middle school students on the effects of global climate change on plants and will provide teachers with tools to teach the quantitative aspects of Ecosystem Health and Sustainability as they relate to plant science.

Sincerely,

A handwritten signature in black ink, reading "Doug Levey". The signature is written in a cursive style with a large, prominent initial "D".

Doug Levey, Ph.D.
Professor and Director of SPICE

PLANT BIOLOGY

Eva Huala, Ph.D.

Adjunct Staff Scientist

huala@acoma.stanford.edu

(650) 325 1521 x310 FAX: (650) 325 6857

SCIENTIFIC DEPARTMENTS

Embryology

BALTIMORE, MARYLAND

Dr. Stephen M Welch
Department of Agronomy
Throckmorton Plant Sciences Center
Kansas State University
Manhattan, KS 66506

Geophysical Laboratory

WASHINGTON, DC

Dr. Ruth Grene
Department of Plant Pathology, Physiology and Weed Science

Global Ecology

STANFORD, CALIFORNIA

103 PMB Building (0330)
Glade Road Research Center
Virginia Tech

The Observatories

PASADENA, CALIFORNIA AND

LAS CAMPANAS, CHILE

Blacksburg, VA 24061

Plant Biology

STANFORD, CALIFORNIA

Dear Steve and Ruth,

I am interested in collaborating with you and the iPlant team to provide cyberinfrastructure support for your projects, “Cyberinfrastructural Support for Genetic and Ecophysiological Studies of Plant Phenological Control in Complex and Changing Environments” and “Cyberinfrastructural Support for Model-Building In Plant Stress Biology: Effects of Climate Change Factors on Crops”.

Terrestrial Magnetism

WASHINGTON, DC

Carnegie Academy for

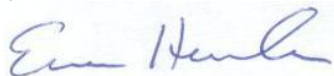
Science Education

WASHINGTON, DC

As director of TAIR I can share my experience and knowledge in the areas described in sections A, B, C and F of your joint Proposed Activities section. TAIR has been very active in biological ontology development and use, data integration and text mining. TAIR/PlantCyc curators also have considerable expertise in curation of metabolic pathways for Arabidopsis and other plant species. In addition since TAIR will presumably serve as a major source of Arabidopsis data for your projects I can help with the integration of this data into the CI.

Depending on the outcome of the TAIR renewal which is currently pending, we may also significantly increase our effort in the area of text mining, and it would be helpful to collaborate in that area.

Sincerely,



Eva Huala, Ph.D

Carnegie Institution
of Washington

260 Panama Street
Stanford, CA 93305

Director, The Arabidopsis Information Resource (TAIR)
Department of Plant Biology
Carnegie Institution of Washington
Stanford, CA

650 325 1521 PHONE

650 325 6857 FAX

White, Jeffrey

From: pstephen.baenziger@gmail.com on behalf of P. Stephen Baenziger [pbaenziger1@unl.edu]
Sent: Saturday, January 24, 2009 3:07 PM
To: White, Jeffrey
Subject: Re: Wheat iPlant group - Y/N response needed by 30 Jan

Jeff:

I will be happy to be part of this collaboration. Are you aware of the National Phenological Network (NPN, <http://www.usanpn.org/>)? I work with them also.

As I read your proposal, your proposal is much more crop and genomics data to simulation modeling oriented than the NPN. NPN is more phenologically and ecosystem based (started long before genomics).

However, NPN is a really superb at the citizen science aspect and is keenly interested in global climate change. I think the iPlant group should contact the NPN as you probably have common interests coming from two different vantage points. It would be good for each group to know of the other's efforts.

Just a thought,

Stephen

On Fri, Jan 23, 2009 at 11:12 AM, White, Jeffrey <Jeffrey.White@ars.usda.gov> wrote:

> We are contacting you to ask whether you would like to collaborate in
> the iPlant Project: "Cyberinfrastructural Support for Genetic and
> Ecophysiological Studies of Plant Phenological Control in Complex and
> Changing Environments." This project seeks to develop a
> "cyberinfrastructure" (CI) that vastly improves the integration of
> data analytic tools across scales ranging from the molecular level to
> field-level phenology of crops and plants in natural ecosystems.
> Emphasis is given to creating models at levels of process detail
> ranging from molecular genetics to field ecophysiology. The CI is
> intended to be flexible and adaptable enough to cover different
> species with minimal adaptation and to be readily extended to other
> traits. To further the goal of broad applicability of tools, this CI
> proposal is being developed in parallel with another one on abiotic stresses.

>

>

>

> Attached please find:

>

> The project summary (one page).

> An executive summary of the full proposal (13 pages).

>

>

>

> The iPlant Collaborative (iplantcollaborative.org) is not designed to
> support extensive experimentation. The assumption is that limitations
> arise from the lack of fluid access to integrated data and support for group work.
> Thus, collaboration should develop through sharing of data resources,
> tools, or test applications or outreach situations.

>

>

> We propose to develop working groups related to various
> aspects of the project. One type of group would be species-related.
> These groups would have the task of identifying sources of molecular, phenotypic and
> environmental data, existing models, and tools. For example, in
> Arabidopsis, there is a vast amount of molecular data available
> relating to the control of floral initiation. In contrast, although
> gene discovery in wheat is constrained by the large genome size and
> lack of efficient means for transformation, the extensive data on
> phenotypes and pedigrees present opportunities for using association
> mapping to identify important loci and inform revision of

Email to:

Dr. Stephen M Welch **or**

Jayme Brown (Sen. Admin Asst.):

Dept Agronomy/KSU

Date: Sat, 24 Jan 2009 17:10:56 +0100

From: George Coupland <coupland@mpiz-koeln.mpg.de>

Reply-To: George Coupland coupland@mpiz-koeln.mpg.de

To: Jayme Brown <jayme2@ksu.edu>

Dear Steve,

thanks for sending your CI proposal and the other attachments.

I would be delighted to take part. We could contribute in the following ways:

- we have extensive flowering time phenotyping data under different environments for Arabidopsis mutants and accessions that we could contribute to the data base and that are likely to contribute to modelling.
- we have quantitative RNA and protein data for several flowering genes under different environments, and intend to extend these. Such data are likely to be very valuable for the molecular modelling.
- I have worked on flowering in Arabidopsis for more than 15 years, and can provide this experience in planning future experiments or in discussing the models.
- Dr Maria von Korff in my department has started a programme on flowering in barley using different mapping populations and transgenic plants. Together with Maria we could contribute on developing crop molecular models.

Please let me know if you need any further information or a more formal declaration of interest.

Best regards

George

White, Jeffrey

From: PAUL GEPTS [plgepts@ucdavis.edu]
Sent: Thursday, January 29, 2009 5:52 PM
To: White, Jeffrey
Subject: RE: Grain legume iPlant group Y/N response needed by 30 Jan

Dear Jeff,

Thank you for getting in touch. I confirm that I am very interested in participating in an iPlant proposal that would involve legumes (and comparisons with other plants).

I believe that there is an outstanding opportunity for a cyberinfrastructure component that would integrate data generated in different legumes, principally of the cool-season (Hologalegoid clade) and the warm-season (Millettoid-Phaseoloid clade). The former clade contains pea, alfalfa, lentil, chickpea, and the "model" legumes *Medicago truncatula* and *Lotus japonicus*. The latter clade contains beans, soybean, cowpea, pigeon pea, winged bean.

The sequences of reference species in the two clades have been or will shortly be published. Soybean is available at: www.phytozome.net/soybean and www.soybase.org. The sequences of Mt and Lj are supposed to come out shortly. Several papers have looked at synteny between the two clades. There seems to be extensive macrosynteny.

From my own, unpublished observations microsynteny is also very high, although not perfect because of translocations, insertions-deletions, etc. in comparisons between soybean and beans. Overlaying phenotypic data on top of synteny maps should be very enlightening. I could provide you with graphics (e.g., screenshots) of soybean-bean comparisons based on the G-browse software.

Let me know if I can be of further assistance,

Paul

From: White, Jeffrey [mailto:Jeffrey.White@ARS.USDA.GOV]
Sent: Friday, January 23, 2009 9:02 AM
To: Boote, Kenneth J; gerrit@uga.edu; Jones, Jimmy W; PAUL GEPTS; S.BEEBE@CGIAR.ORG; m.blair@CGIAR.ORG; mbrick@lamar.colostate.edu; mgmea@ifas.ufl.edu; j_beaver@rumac.upr.clu.edu
Cc: Stephen Welch; correllm@ufl.edu
Subject: Grain legume iPlant group Y/N response needed by 30 Jan

We are contacting you to ask whether you would like to collaborate in the iPlant Project: "Cyberinfrastructural Support for Genetic and Ecophysiological Studies of Plant Phenological Control in Complex and Changing Environments." This project seeks to develop a "cyberinfrastructure" (CI) that vastly improves the integration of data analytic tools across scales ranging from the molecular level to field-level phenology of crops and plants in natural ecosystems. Emphasis is given to creating models at levels of process detail ranging from molecular genetics to field ecophysiology. The CI is intended to be flexible and adaptable enough to cover different species with minimal adaptation and to be readily extended to other

01/30/2009

Email to:

Dr. Stephen M Welch or
Jayme Brown (Sen. Admin Asst.):
Dept Agronomy/KSU
Dear Steve:

I will be happy to collaborate with your group on the iPlant GC project entitled "Cyberinfrastructural Support for Genetic and Ecophysiological Studies of Plant Phenological Control in Complex and Changing Environments". I would like to participate in the grain legume (GL) working group, especially on soybean genomics-related aspects. We have maturity data in hand from a large Recombinant Inbred Line (RIL) soybean population. The RILs have been genotyped with a large number of molecular markers which we have used for QTL mapping of various traits including maturity (flowering time). We also have gene expression data on each line using Affymetrix Soybean GeneChips.

Please let me know if you need any further information from me at this stage. Thank you for the opportunity to participate in this exciting GC project. My address info is at the end of this message.

Regards,

Saghai

Saghai Maroof
Professor
Dept. of Crop and Soil Environmental Sciences
Virginia Tech
Blacksburg, VA 24060

Phone: 540-231-9791
Email: smarroof@vt.edu

Email to:

Dr. Stephen M Welch **or**
Jayme Brown (Sen. Admin Asst.):
Dept Agronomy/KSU
Dear Steve,

I apologize for the late response. In collaboration with Justin Borevitz and Joy Bergelson, we are currently genotyping over 1,000 accessions of *A. thaliana* for 250,000 SNPs. We have also genotypes over 6,000 accessions using 149 SNPs as a form of fingerprinting. We are working on integrating all available polymorphism data for *A. thaliana*, including the rapidly accumulating whole-genome sequences.

In addition, we are carrying out genome-wide association mapping, and have accumulated over 100 phenotypes for a subset of 192 genotyped accessions.

A long-term goal is the establishment of web-site that gives the community access to these kinds of data. It would be very sensible to coordinate efforts with your proposed project so duplication of effort is avoided.

Would be delighted to discuss further, when/if needed.

Please let me know if you need some kind of more formal letter.

Cheers,

Magnus

Magnus Nordborg, Scientific Director
Gregor Mendel Institute
Dr. Bohr-Gasse 3
1030 Vienna, Austria
T: +43 1 79044-9000
F: +43 1 79044-9001
magnus.nordborg@gmi.oeaw.ac.at

Email to:

Dr. Stephen M Welch or
Jayme Brown (Sen. Admin Asst.):
Dept Agronomy/KSU

Date: Fri, 23 Jan 2009 20:50:22 +0100

From: Detlef Weigel <weigel@weigelworld.org>

Reply-To: Detlef Weigel <weigel@weigelworld.org>

Dear Stephen,

there wasn't anything attached, but we are certainly interested in contributing to this. We are generating comprehensive genome sequence data of *A. thaliana*, with the goal to sequence 1001 genomes over the next two years (<http://1001genomes.org>). We are also producing extensive phenotype data, and have assembled local collections of *A. thaliana* that have been genotyped at hundreds of markers, and many of which will be completely genotyped.

Cheers

Detlef Weigel
Max Planck Institute for Developmental Biology

Email to:

Dr. Stephen M Welch or
Jayme Brown (Sen. Admin Asst.):
Dept Agronomy/KSU
Dear Steve,

I'm sorry for not responding earlier, but I did get a chance to look over the summary, and it looks very exciting.

I certainly would like to be involved as a participant in all the ways you outline below. While *Mimulus* is not as far along as *Arabidopsis*, we are doing an increasing amount of work on genetic variation in flowering responses to photoperiod, vernalization, etc. and are beginning to clone QTLs for these traits. The *Mimulus* community certainly could make their data available and would also like to test run the software. And using the educational materials in our own teaching and outreach would also be great.

Please write me in, and let me know what I need to write if anything. I have some time next week to work on it if needed.

Cheers, and many thanks for being persistent - - I'm a terrible manager of my inbox!!!
John

John H. Willis
Department of Biology; Box 90338
3314 French Family Science Center
Science Drive
Duke University
Durham, NC 27708

office phone: (919) 660-7340

jwillis@duke.edu

Email to:

Dr. Stephen M Welch or
Jayme Brown (Sen. Admin Asst.):
Dept Agronomy/KSU

Date: Sat, 31 Jan 2009 10:44:19 +0800
From: Xifeng Yan <xifeng@gmail.com>
To: welchsm@ksu.edu

Hi Welch,

It was nice to chat with you in the Nov meeting. I joined the CS department at Univ. of California at Santa Barbara a few weeks ago. Here is my web link, <http://www.cs.ucsb.edu/~xyan>. I also hold the Venkatesh Narayanamurti chair there . My research interests include graph/network mining, modeling, and management. I am interested to supply my algorithms to integrate, search, and mine biological graphs and networks that might be used in your project. In the past 5 years, I have developed a full range of graph mining and management tools that are able to query graphs, find patterns, compare graphs, etc. Please let me know if it is interesting to your project. If so, I will be very glad to join. By the way, we have a strong network visualization team here, if needed, I could help get you connected. Thanks.

Best Regards,
Xifeng

Email to:

Dr. Stephen M Welch or
Jayme Brown (Sen. Admin Asst.):
Dept Agronomy/KSU
Steve,

I'm sorry for not responding. I am interested in all 3 areas-- I do try to make it a point to format and release data for others to interrogate; I have interest in test-driving and advising-- I do this for PLEXdb along with module design/initiation; I do not teach (I have an ARS appointment), but am willing to share seminar presentations.

The entire USDA email was closed by the FBI a few weeks ago. I just got my access back. We now must VPN to central servers for email, so I'd like to continue the iPlant dialoge via nickl@iastate.edu

Please feel free to call me-- this might be a quick way for us to see where and how I could help.

Again, sorry for the communication shortage to this point.

Nick

Nick Lauter
USDA-ARS Research Geneticist
Corn Insects and Crop Genetics Research Unit
Department of Plant Pathology and Microbiology
415 Bessey Hall
Iowa State University, Ames, IA
50011-1020 USA

Office: 515-294-8260
Lab: 515-294-7577
Fax: 515-294-9420
E-mail: nick.lauter@ars.usda.gov

White, Jeffrey

From: g.mclaren@cggmail.org on behalf of Graham McLaren (IRRI-CIMMYT) [g.mclaren@cgiar.org]
Sent: Monday, February 02, 2009 11:55 AM
To: welchsm@ksu.edu
Cc: Bruskiwich, Richard (IRRI); Metz, Thomas (IRRI); Davenport, Guy (CIMMYT); White, Jeffrey; Senger, Martin (IRRI); tpb8@cornell.edu
Subject: Re: iPlant phenology project
 Steve, Richard, Jeff and colleagues,

Firstly my sincere apologies for being so slow to join this debate, and my thanks for Richard for putting it into context in his usual visionary way.

Clearly we would love to see ICIS linked into the iPlant initiative. Richard has pointed out what that could involve - public Crop Information for several crops including Rice Wheat and Maize, a set of tools for curating such information at the field level and several steps on the way to a next generation of web based applications.

As Ricard mentioned, and my excuse for lack of engagement, is that I have been putting together a Gates proposal to develop a Molecular Breeding Platform. One element of that is an Information System, and ICIS is squarely at the center of it. The question is how does this link with iPlant. I did not detect a readiness on the part of the iPlant community to concentrate on what we call Crop Information - basically the genetic resource evaluation and improvement spectrum. Is this true? Or is there some element of iPlant emerging to deal with this. If my observation is correct, we have opportunities for synergy and I hope we can get them through projects like the Phenology CP.

I would very much like to remain in touch with iPlant and my role would be to try and maintain the linkages between GCP activities and iPlant activities in the area of Crop Information Systems. Clearly one of the first issues raised will be the integration of the database you mention TAIR, PlantGDB, ICIS, GRIN etc. This is explicitly the objective of the GCP Informatics Platform project. Similarly we have a lot of work going on to develop plant and trait ontologies for several focus crops. The Molecular Breeding Platform has a large component on modelling, both quantitative genetics and eco-physiological (and their integration). However the emphasis of these tasks will be on trait prediction to help breeding rather than understanding physiological mechanisms (ie we want to deploy that understanding developed by the phenology CP as Richard remarked).

IF these contributions and connections are enough for me to bring to the table, I would certainly like to be part of this CI.

Regards,
 Graham

On Mon, Feb 2, 2009 at 8:42 AM, <welchsm@ksu.edu> wrote:

Hi Richard:

Let me add my comments to what you have from Jeff under separate cover.

After receiving your email below, I spent some time reviewing materials at the pantheon website and, as always, due to the care and expertise you and your team bring to bear, they were quite impressive. I think

Email to:

Dr. Stephen M Welch or
Jayme Brown (Sen. Admin Asst.):
Dept Agronomy/KSU

Date: Fri, 23 Jan 2009 14:01:34 -0500

From: sheldong1@aol.com

Reply-To: sheldong1@aol.com

Hi Jayme,

Thank you for the note regarding the iPlant Phenology Project.? I would be delighted to participate in the Education and Outreach portion of this project.? I noticed that Appendix 3 lists my expertise as access to large numbers of active citizen scientists.? I should also mention that I am the Chief Academic Officer of an online university and therefore have considerable experience in the development and deployment of distance learning materials meeting national accreditation standards.? I would be happy to apply that expertise to the development or critique of any educational materials developed for this project.

One other item I would like to mention is that SAS has a web site that is nearing completion.? You might want to consider this as the basis for your CI, since it allows for the kind of on-the-fly adjustments in data collection and changes your program calls for.? Additionally, the site is equipped with a suite of statistical tools and plots all observations on a Google map.? A cursory reading of your proposal leads me to think that our site would serve the purposes of your project quite well. If you wish, I will convey the proposal to our IT Director to see if he agrees that our site would work for you.

Thank you once again for the invitation.? I look forward to working with you and your colleagues.

Best wishes,

Sheldon

White, Jeffrey

From: Guisard, Yann [yguisard@csu.edu.au]
Sent: Tuesday, January 27, 2009 9:00 PM
To: White, Jeffrey
Subject: RE: Collaboration in ecophysiological modeling -- Y/N response needed by 30 Jan
Attachments: Dr Yann Guisard.vcf

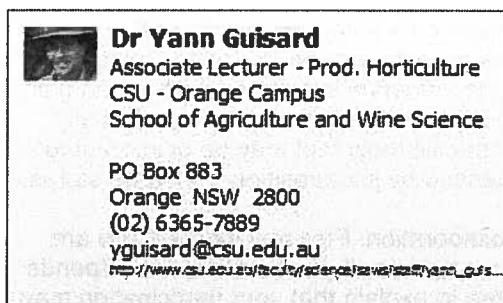
Dear Sir,

I would like to express my interest in participating in the in the iPlant Project: "Cyberinfrastructural Support for Genetic and Ecophysiological Studies of Plant Phenological Control in Complex and Changing Environments."

At this stage, my participation is dependent upon the level of travel (mostly from a cost perspective) associated with the proposed model of collaboration.

My particular interest is in three dimensional modelling of biophysical processes in orchards and vineyards, with particular interest in stomatal behaviour, predictive growth and visualisation. Field of application include tertiary education and the design of targeted irrigation strategies. Although my principal crops of interest are perennial plants, I am happy to participate in the development of process based modelling of other plants.

Many thanks



YOU MUST READ THIS NOTICE

This email has been sent by Charles Sturt University (CSU) (ABN 83 878 708 551, CRICOS 00005F). This email (and any attachment) is confidential and is intended for the use of the addressee(s) only. If you are not the intended recipient of this email, you must not copy, distribute, take any action in reliance on it or disclose it to anyone. Any confidentiality is not waived or lost by reason of mistaken delivery. The views expressed in this email are not necessarily those of CSU. Email should be checked for viruses and defects before opening. CSU does not accept liability for viruses or any consequence which arise as a result of this email transmission. Email communications with CSU may be subject to automated email filtering, which could result in the delay or deletion of a legitimate email before it is read at CSU.

From: ICASA Members [mailto:ICASA@listserv.uga.edu] **On Behalf Of** White, Jeffrey
Sent: Saturday, 24 January 2009 7:09 AM

01/30/2009

White, Jeffrey

From: Johannes Muller [johannes.mueller@landw.uni-halle.de]
Sent: Monday, January 26, 2009 9:34 AM
To: White, Jeffrey
Subject: AW: Collaboration in ecophysiological modeling -- Y/N response needed by 30 Jan

Dear Professor White,

in my opinion, this is a very interesting proposal that fits well to the main focus of my research. So, I am interested to participate, in particular concerning the following fields:

- interactions between processes of plant development and formation of dry mass and yield,
- integration of these interactions into complex crop models on a more sound basis,
- links to genomics (barley)
- specification of leaf and canopy gas exchange models, in particular with respect of the effects of plant development and senescence.

I strongly recommend to include barley by reason of its worldwide importance in agriculture and as a model plant in plant genetics/genomics and plant breeding. Barley is the model plant of several collaborative research networks focussing on items that are close to your proposal such that there may result synergies.

Some funding would be required for a substantial contribution.

Please send the executive summary of the full proposal.

Kind regards

Johannes

Johannes Müller, PhD
Associate Professor
Martin-Luther-University Halle-Wittenberg
Institute of Agricultural and Nutritional Sciences
Dep. of Crop Science and Yield Physiology
Group Crop and Ecological Physiology / Plant Systems Modelling
Ludwig-Wucherer Straße 2
D-06108 Halle/Saale
Germany
Fon: ++49 345 5522 605
Fax: ++49 345 5527 119
www.landw.uni-halle.de/prof/spb

Email to:

Dr. Stephen M Welch or
Jayme Brown (Sen. Admin Asst.):
Dept Agronomy/KSU

Date: Sat, 24 Jan 2009 11:16:10 +0800

From: Xinguang Zhu <zhuxinguang@picb.ac.cn>

Reply-To: Xinguang Zhu <zhuxinguang@picb.ac.cn>

Dear Steve and Jeff,

Though I am now working with Tom Brutnell from Cornell, Richard from IRRI to submit a proposal intended for photosynthesis modeling, targeting at the C4 engineering, I would be happy to participate this iPlant project.

I would be able to contribute to this project the detailed mechanistic models of photosynthesis (already published) and models of the plant primary metabolism that will be developed from my group. The mechanistic models can directly link the proteomics data and metabolomics data to leaf level physiological data.

Now, it is Chinese New Year Period. I would like to extend my best wishes to you as well.

My formal address now:

Junior Group Leader, Plant Systems Biology Group,
Partner Institute of Computational Biology, Chinese Academy of
Sciences/Max
Planck Society
Shanghai, China, 200031

Best,

Xinguang

JUSTIN O. BOREVITZ

PROFESSIONAL PREPARATION

University of California, Santa Cruz	Biochemistry & Molecular Biology	B.S.	1997
University of California, San Diego	Biology	Ph.D.	2002
The Salk Institute for Biological Studies	Genomics		2002-2004

APPOINTMENTS

2004-present Assistant Professor in Dept. of Ecology & Evolution, University of Chicago
2002-2004 Helen Hay Whitney Postdoctoral Fellow, Genomics, The Salk Institute for Biological Studies

PUBLICATIONS

5 Most Closely Related to Proposed Project

- Shiu SH, Borevitz JO. The next generation of microarray research: applications in evolutionary and ecological genomics. *Heredity*. 2008 Feb;100(2):141-9. Epub 2006 Nov 8. PMID: 17091126
- Zhang X, Shiu S, Cal A, Borevitz JO. Global analysis of genetic, epigenetic and transcriptional polymorphisms in *Arabidopsis thaliana* using whole genome tiling arrays. *PLoS Genet*. 2008 Mar 21;4(3):e1000032. PMID: 18369451.
- Borevitz JO, Hazen SP, Michael TP, Morris GP, Baxter IR, Hu TT, Chen H, Werner JD, Nordborg M, Salt DE, Kay SA, Chory J, Weigel D, Jones JD, Ecker JR. Genome-wide patterns of single-feature polymorphism in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A*. 2007 Jul 17;104(29):12057-62. Epub 2007 Jul 12.
- Li Y, Roycewicz P, Smith E, Borevitz JO. Genetics of local adaptation in the laboratory: flowering time quantitative trait loci under geographic and seasonal conditions in *Arabidopsis*. *PLoS ONE*. 2006 Dec 27;1:e105.
- Werner JD, Borevitz JO, Warthmann N, Trainer GT, Ecker JR, Chory J, Weigel D. Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. *Proc Natl Acad Sci U S A*. 2005 Feb 15;102(7):2460-5. Epub 2005 Feb 4. PMID: 15695584

5 Additional Publications

- Emerson JJ, Cardoso-Moreira M, Borevitz JO, Long M. Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science*. 2008 Jun 20;320(5883):1629-31. Epub 2008 Jun 5. PMID: 18535209.
- Filialt DL, Wessinger CA, Dinneny JR, Lutes J, Borevitz JO, Weigel D, Chory J, Maloof JN. Amino acid polymorphisms in *Arabidopsis phytochrome B* cause differential responses to light. *Proc Natl Acad Sci U S A*. 2008 Feb 26;105(8):3157-62. Epub 2008 Feb 14. PMID: 18287016.
- Jones-Rhoades MW, Borevitz JO, Preuss D. Genome-wide expression profiling of the *Arabidopsis* female gametophyte identifies families of small, secreted proteins. *PLoS Genet*. 2007 Oct 12;3(10):1848-61. Epub 2007 Aug 24.

Borevitz, J.O., Maloof, J.N., Lutes, J., Dabi, T., Redfern, J.L., Trainer, G.T., Werner, J.D., Asami, T., Berry, C.C., Weigel, D., and Chory, J. 2002. Quantitative Trait Loci Controlling Light and Hormone Response in Two Accessions of *Arabidopsis thaliana*. *Genetics* 160(2):683-96.

Rostoks N, Borevitz JO, Hedley PE, Russell J, Mudie S, Morris J, Cardle L, Marshall DF, Waugh R. Single-feature polymorphism discovery in the barley transcriptome. *Genome Biol.* 2005;6(6):R54. Epub 2005 May 11.

SYNERGISTIC ACTIVITIES

Establishing a Center for Ecological Conservation Genetics with the Field Museum and UChicago.

Developed a field project undergraduate course on Prairie Ecosystems
<http://natural.uchicago.edu/prairieecosystems>

Developing methods for haplotype identification from high resolution SFP data

Assisted on the development of BQTL (Bayesian QTL mapping) and used BQTL to perform two-dimensional QTL epistasis scans using permutation thresholds (2001)

American Society of Plant Biologists and American Association for the Advancement of Science (1998-Present)

COLLABORATORS & OTHER AFFILIATIONS

Collaborators and Co-Editors

Magnus Nordborg, University of Southern California; Paul Marjoram, University of Southern California; Sebastian Zoellner, University of Michigan; Scott Hodges, University of California-Santa Barbara; Elena Kramer, Harvard University; Jeff Tomkins, Clemson University

Graduate and Postdoctoral Advisors

(Graduate) Joanne Chory Salk Institute for Biological Studies and Howard Hughes Medical Institute
(Postdoctoral) Joseph Ecker Salk Institute for Biological Studies

Thesis Advisor and Postdoctoral-Scholar Sponsor

Dissertation Directed

Current:

Graduate Student (1 total)

Andy Cal Department of Molecular Genetics and Cell Biology

Postdoctoral fellows (4 total)

Former:

None

Current:

Xu Zhang Postdoctoral Scholar

Yan Li Postdoctoral Scholar

Christos Nostos Postdoctoral Scholar

Geoff Morris Postdoctoral Scholar

Edward S. Buckler

Institute for Genomic Diversity
159 Biotechnology Bldg.
Cornell University
Ithaca, NY 14850-2703

Email: esb33@cornell.edu
Voice: (607) 255-4520
Fax: (607) 255-6249
<http://www.maizegenetics.net/>

A. Education:

University of Virginia	Biology and Archaeology	BA(w/ Honors)	1992
University of Missouri	Biological Sciences	PhD	1997
North Carolina State Univ.	Statistical Genetics	Postdoc	1997-1998

B. Appointments:

1998-Present USDA-ARS Research Geneticist
1998-2003 Assistant Professor, Department of Genetics, North Carolina State University
2003-Present Adjunct Associate Professor, Plant Breeding, Cornell University

C. Publications Related:

Thornsberry, J.M., M.M. Goodman, J. Doebley, S. Kresovich, D. Nielsen, and E. S. Buckler. 2001. *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics*. **28**: 286-289.

Remington, D. L., J. Thornsberry, Y. Matsuoka, L. Wilson, S. Rinehart-Whitt, J. Doebley, S. Kresovich, M. M. Goodman, and E. S. Buckler. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Nat. Acad. Sci. USA* **98**:11479-11484.

Yu, J., Pressoir, G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Holland, J.B., Kresovich, S., and Buckler, E.S. 2006. A Unified Mixed-Model Method for Association Mapping Accounting for Multiple Levels of Relatedness. *Nature Genetics* **38**:203-208.

Flint-Garcia, S.A., A.-C. Thuillet, J.M. Yu, G. Pressoir, S.M. Romero, S.E. Mitchell, J. Doebley, S. Kresovich, M.M. Goodman, and E.S. Buckler. 2005. Maize association population: a high resolution platform for QTL dissection. *Plant Journal* **44**: 1054-1064.

Harjes, C.E., Rocheford, T., Sowinski, S., Williams, M., Kandanis, C.B., Brutnell, T., Wurtzel, E., Yan, J., and Buckler, E.S. 2008. Natural genetic variation in lycopene epsilon cyclase can enhance provitamin a biofortification of maize. *Science* **319**: 330-333.

Other Publications:

Whitt, S. R., Wilson, L. M., Tenailon, M. I., Gaut, B. S. & Buckler, E. S. 2002. Genetic diversity and selection in the maize starch pathway. *Proc. Nat. Acad. Sci. USA* **99**: 12959-12962.

Wilson, L.M., Whitt, S.R., Ibanez-Carranza, A.M., Goodman, M.M., Rocheford, T.R., and Buckler, E.S. 2004. Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell* **16**: 2719-2733.

Viviane Jaenicke, E. S. Buckler, Bruce D. Smith, M. Thomas P. Gilbert, Alan Cooper, John Doebley, Svante Pääbo. 2003. Early allelic selection in maize as revealed by ancient DNA. *Science* **302**: 1206-1208.

Stich B, Yu J, Hans P. Maurer, Melchinger AE, Utz HF, Buckler ES 2007. Power to Detect Higher-Order Epistatic Interactions in a Metabolic Pathway Using Nested Association Mapping and Diallel Association Mapping *Genetics* **176**: 563-570

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*. **23**: 2633-2635.

D. Synergistic Activities:

1. Developing database aware software and analysis approaches for association and joint linkage-association mapping. Currently, TASSEL analyses molecular diversity and carries out a wide range of association and molecular quantitative genetic analyses.
2. Involved in designing the bioinformatics architecture for sharing and storing diversity data. As a co-PI on the Gramene project, we have developed a diversity schema, middleware, and query tools to share basic diversity data (both genotypic and phenotypic).

E. Collaboration and Other Affiliations

i. Collaborators in past 48 months

S. Avraham (Cold Spring Harbor), L. Bai (Boyce Thompson Inst.), C. Bermudez Kandanis (U. Illinois), I. Bi Vroh (IITA), P. Bradbury (USDA-ARS), W. Briggs (U. of Wisconsin), T. Bruhnell (Boyce Thompson Inst.), R. Caldo (Iowa State Univ.), P. Canaran (Cold Spring Harbor), P. Casati (Stanford), T. Casstevens (Cornell U.), J. Doebley (U. Wisconsin), E. Ersoz (Cornell U.), E. Ersoz, Elhan (Cornell U.), E. Esch (U. Hannover), B. Faga (Cold Spring Harbor), S. Flint-Garcia (North Carolina State U.), M. Fogleman (Cornell U.), K. Fukunaga (U. of Wisconsin), T. Fulton (Cornell U.), B. Gaut (U. California, Irvine), J. Glaubitz (U. of Wisconsin), L. Goh (Cold Spring Harbor), M. Goodman (North Carolina State U.), M. Gore (Cornell U.), C. Harjes (Monsanto), C. Hebbard (Cornell U.), J. Heer (North Carolina State U.), J. Hill (U. of Wisconsin), O. Hoekenga (USDA/Cornell U.), R. Hogers (Keygene), J. Holland (USDA/N. Carolina State U.), T. Holtsford (U. Missouri), P. Jaiswal (Cornell U.), R. Jung (Pioneer Hi-Bred), R. Jurkuta (U. of Wisconsin), M. Kirst (U. of Florida), L. Kochian (USDA/Cornell U.), S. Kresovich (Cornell U.), D. Kroon (Cornell U.), C. Liang (Cold Spring Harbor), K. Liu (North Carolina State Univ.), J. Maohringy (U. Hohenheim), R. Martienssen (Cold Spring Harbor), K. Matias (U. Florida), Y. Matsuoka (U. of Wisconsin), H. Maurer (U. Hohenheim), S. McCouch (Cornell U.), M. McMullen (Univ. of Missouri), A. Melchinger (U. Hohenheim), S. Mitchell (Cornell U.), J. Ni (Cornell U.), D. Nielsen (North Carolina State U.), W. Pawlowski (Cornell U.), J. Peleman (Keygene), H. Piepho (U. Hohenheim), G. Pressoir (Cornell U.), Y. Ramdoss (Cisco), K. Ratnapu (Cold Spring Harbor), L. Ren (Cold Spring Harbor), T. Rocheford (U. Illinois), S. Romero (Cornell U.), J. Rose (Cornell U.), S. Schmidt (Cold Spring Harbor), M. Snook (USDA-ARS), S. Sowinski (Pioneer Hi-Bred), W. Spooner (Cold Spring Harbor), P. Springer (Cold Spring Harbor), A. Stapleton (U. North Carolina, Wilmington), L. Stein (Cold Spring Harbor), N. Stevens (Cornell U.), B. Stich (U. Hohenheim), Q. Sun (Cornell U.), S. Szalma (Univ. of Missouri), J. Szymaniak (Cornell U.), G. Tanimoto (Affymetrix), A. Thuillet (U. of Wisconsin), R. Toddhunter (Cornell U.), W. Tracy (U. of Wisconsin), F. Utz (U. Hohenheim), R. Vallabhaneni (CUNY), M. van Eijk (Keygene), J. van Oeveren (Keygene), E. Verstege (Keygene), Y. Vigouroux (U. of Wisconsin), D. Vleck (USDA-ARS (Lincoln)), E. Vollbrecht (Iowa State Univ.), V. Walbot (Stanford), D. Ware (Cold Spring Harbor), S. Whitt (North Carolina State U.), M. Williams (Pioneer Hi-Bred), R. Wise (Iowa State Univ.), E. Wurtzel (CUNY), L. Yamasaki (U. of Missouri), J. Yan (CIMMYT), I. Yap (Cornell U.), H. Yates (Cornell U.), K. Youens-Clark (Cold Spring Harbor), J. Yu (Cornell U.), Z. Zhang (Cornell U.), W. Zhao (Cold Spring Harbor)

ii. Graduate and Postdoctoral Advisors

T. P. Holtsford (U. Missouri)

B. S. Weir (North Carolina State U.)

M. Purugganan (North Carolina State U.)

iii. Thesis advisor and postdoctoral sponsor

Beth K. Saft, Sherry Flint-Garcia, Jeff M. Thornsberry, David M. Remington, Larissa M. Wilson (North Carolina State U); Elhan Ersoz, Amit Gur, Michael Gore, Carlos Harjes, Allison Krill, Matias Kirst, Jason Pfeiffer, Feng Teng, Irie Vroh, Jianming Yu, Nengyi Zhang (Cornell Univ.)

Biographical Sketch: Melanie J. Correll

Assistant Professor
Department of Agricultural and Biological Engineering
University of Florida
P.O. Box 110570
Gainesville, FL 32611-0570
Ph: (352) 392-1864 ext. 209; E-mail: correllm@ufl.edu

Education:

Clemson University, B.S., Agricultural and Biological Engineering, 1994
Worcester Polytechnic Institute, Ph.D., Interdisciplinary in Biochemical Engineering/Plant Science, 2001
Miami University, Postdoctoral Associate, Plant Physiology, 2001–2004

Appointments:

2004-present Assistant Professor, Agricultural and Biological Engineering Department, Univ. of Florida

Publications (5 publications most closely related to proposed project):

- Stimpson, A.J., R.S. Pereira, J.Z. Kiss, and M.J. Correll. Extraction and labeling methods for microarrays using small amounts of plant tissue. *Physiol Plant*. In press.
- Molas, M.L., J.Z. Kiss, and M.J. Correll. 2006. Gene Profiling the Red-Light Signaling Pathways In Roots. *J Exp Bot*. 57: 3217-3229.
- Correll, M.J. and J.Z. Kiss. 2005. The roles of phytochromes in elongation and gravitropism of roots. *Plant Cell Physiol*. 46:317-323.
- Correll, M.J., K.M. Coveney, S.V. Raines, J.L. Mullen, R.P. Hangarter, and J.Z. Kiss. 2003. Phytochromes play a role in phototropism and gravitropism in Arabidopsis roots. *Adv. Space Res*. 31:2203-2210
- Kiss, J.Z., J.L. Mullen, M.J. Correll, and R.P. Hangarter. 2003. Phytochromes A and B mediate red-light-induced positive phototropism in roots. *Plant Physiol*. 131:1411-1417.

Publications (5 other publications):

- Lee, S-C., W.-C. Hu, M.J. Correll, and J. Xin. 2006. Accessing genome databases from internet-enabled mobile handheld devices. *The Proceedings for the 4th World Congress on Computers in Agriculture*.
- Correll, M.J. and J.Z. Kiss. 2002. Interactions between gravitropism and phototropism in plants. *J. Plant Growth Reg*. 21:89-101.
- Kiss, J.Z., M.J. Correll, J.L. Mullen, R.P. Hangarter, and R.E. Edelman. 2003. Root phototropism: how light and gravity interact in shaping plant form. *Gravitational and Space Biology Bulletin* 17: 16:55-60.
- Correll, M.J. and P.J. Weathers. 2001. One step acclimatization of plantlets using a mist reactor. *Biotech. Bioeng*. 73(3):253-258.
- Correll, M.J. and P.J. Weathers. 2001. Effects of light, CO₂, and humidity on carnation growth, hyperhydration, and cuticular wax development in a mist reactor. *In Vitro Cell. Dev. Biol.-Plant*. 37(3):405-413.

Awards and Honors

- 2007 Appreciation Award for Outstanding Participation on the TROPI experiment, from European Space Agency and National Aeronautics and Space Administration
- 2007 Young Research Award, Florida Section of the American Society of Agricultural and Biological Engineers
- 1990-1994 USDA National Needs Fellowship

Synergistic Activities

(a) Editor, *Advances in Space Research* Volume 39, Issue 7, 2007, (b) Reviewer for scientific Journals (*Advances in Space Research*, *Journal of Biological Engineering*, *In Vitro-Plant*, *American Society of Space and Gravitational Biology Bulletin*) (c) Mentor for high school students in the Student Science Training Program (2006, 2007), (d) 4-H workshop development and presentation on Space Agriculture (2008), (e) Development of a new undergraduate course and labs entitled “Applications in Biological Engineering” (f) Session chair Congress on Space Research (COSPAR 2008), (g) Publications committee member (Institute of Biological Engineers) (h) Speaker and mentor for the Center for Precollegiate Education on Biological Engineering for recruitment of minorities (i) Science for Life mentor for undergraduate research.

Recent Collaborations: Robert Bowmann (NASA-Ames), Ray Bucklin (Univ of FL), Khe Chau (Univ. of FL), Joe H. Cho (Univ. Central FL), Alex DiIorio (Worcester Polytechnic Institute), Richard Edelmann (Miami Univ.), Mike Eoidice (NASA-Ames), Robert Ferl (Univ. of FL), Bin Gao (Univ. of FL), Maria Gallo (Univ. of Florida), Ruth Grene (Virginia Tech.), Roger P. Hangarter (Indiana Univ.), Jack L. Mullen (Colorado State), M. Lia Molas, (Argentina Dept. Agriculture) Quinn Li (Miami Univ.), Pamela Ronald (UC Davis), Sudipta Seal (Univ. Central FL), Marianne Steele (NASA-Ames), Steve Welch (Kansas State University), Jeff White (USDA-ARS, Tempe), Barbara Wyslouzil (The Ohio State Univ.)

Graduate and Post Doctoral Advisors:

Pamela J. Weathers, Ph.D. Advisor, Worcester Polytechnic Institute, MA and postdoctoral advisor John Z. Kiss, Miami University, OH.

Graduate Students:

Thesis advisor (Chair): Ph.D. Current: Richard (S.C.) Lee, Hemant Gohil, Former: M.S. student Yibing Fu. Undergraduates: Current (three); Former (eight).

Thesis advisor (committee): Ph.D. Abbay Koppar, Sachin Gadekar.

Chad L. Myers, Ph.D.

Department of Computer Science and Engineering, 200 Union Street SE, University of Minnesota
Minneapolis, MN 55455

Phone: (612) 624-8306, Fax: (612) 625-0572

E-mail: cmyers@cs.umn.edu

<http://www.cs.umn.edu/~cmyers>

Professional Preparation

- Ph.D. 2007 Computer Science, Princeton University, NJ
Dissertation: Context-sensitive Methods for Learning from Genomic Data
- B.S. 2002 Computer Engineering and Mathematics, Southern Methodist University, TX
Summa cum laude, Honors in Liberal Arts
- B.A. 2002 Physics, Southern Methodist University, TX

Positions

- 2008- Assistant Professor, Dept. of Computer Science and Engineering, University of
Minnesota, MN
- 2002-07 Ph.D. Student, Dept. of Computer Science and Lewis-Sigler Institute for Integrative
Genomics, Princeton University, NJ
- 2000-01 Undergraduate Research Assistant, Department of Electrical Engineering, Southern
Methodist University, TX

Selected Publications

Most relevant:

Myers CL, Troyanskaya OG: **Context-sensitive data integration and prediction of biological networks**. *Bioinformatics* 2007, **23**(17):2322-2330.

Myers CL, Barrett DR, Hibbs MA, Huttenhower C, Troyanskaya OG: **Finding function: evaluation methods for functional genomic data**. *BMC Genomics* 2006, **7**:187.

Myers CL, Robson D, Wible A, Hibbs MA, Chiriac C, Theesfeld CL, Dolinski K, Troyanskaya OG: **Discovery of biological networks from diverse functional genomic data**. *Genome Biol* 2005, **6**(13):R114.

Guan Y, Myers CL, Lu R, Lemischka IR, Bult CJ, Troyanskaya OG: **A genome-wide functional network for the laboratory mouse**. *PLoS Computational Biology* (*in press*) 2008.

Hibbs MA, Myers CL, Huttenhower C, Hess DC, Li K, Caudy AA, Troyanskaya OG: **Analysis of Computational Functional Genomic Approaches for Directing Experimental Biology**. *PLoS Computational Biology* (*in press*) 2008.

Other selected:

Myers CL, Chen X, Troyanskaya OG: **Visualization-based discovery and analysis of genomic aberrations in microarray data.** *BMC Bioinformatics* 2005, **6**(1):146.

Myers CL, Dunham MJ, Kung SY, Troyanskaya OG: **Accurate detection of aneuploidies in array CGH and gene expression microarray data.** *Bioinformatics* 2004, **20**(18):3533-3543.

Guan Y, Myers CL, Hess DC, Barutcuoglu Z, Caudy AA, Troyanskaya OG: **Predicting gene function in a hierarchical context with an ensemble of classifiers.** *Genome Biol* 2008, **9 Suppl 1**:S3.

Hibbs MA, Hess DC, Myers CL, Huttenhower C, Li K, Troyanskaya OG: **Exploring the functional landscape of gene expression: directed search of large microarray compendia.** *Bioinformatics* 2007.

Miller DL, Myers C, Rickards B, Collier H, Flint J: **Adenovirus type 5 exerts genome-wide control over cellular programs governing proliferation, quiescence and survival.** *Genome Biol* 2007, **8**(4):R58.

Synergistic Activities

1. Development of methods and visualization software for genomic data integration and prediction of biological networks. I have developed or co-developed six different software packages for genomic data processing or integration, all of which are freely available as desktop applications or web-based interfaces. Several of these are integrated with the *Saccharomyces* Genome Database (SGD) and used by the yeast community on a daily basis.
2. Course development for "Introduction to Perl and Matlab for Biology" co-taught with K.C. Huang, and Greg Reeves. We voluntarily developed and taught this course for graduate students and post-docs in molecular biology to teach them the fundamentals of programming in both Perl and Matlab. The course material, including a number of hands-on problems with example scripts, are publicly available on the web.
3. Course development for "Functional Genomics, Systems Biology and Bioinformatics" co-taught with Rui Kuang at the University of Minnesota. This introductory graduate-level course covers computational approaches in functional genomics including gene expression analysis, analysis and prediction of protein-protein interaction maps, and data integration.

Biographical sketch: Johanna Schmitt

A. Professional preparation

Swarthmore College	Biology	B.A. with distinction 1974
Stanford University	Biological Sciences	Ph.D. 1981
Duke University	Ecological genetics	Postdoc 1981-1982

B. Appointments

2003-	Stephen T. Olney Professor of Natural History, Brown University
2001-2003	Harrison S. Kravis University Professor, Brown University
2000-	Professor of Environmental Studies, Brown University
1994-	Professor of Biology, Brown University
1987-1994	Associate Professor of Biology, Brown University
1982-1987	Assistant Professor of Biology, Brown University

C. Publications

i. Five publications most closely related to the current project:

- Weinig, C., M.C. Ungerer, L.A. Dorn, N.C. Kane, S.S. Halldorsdottir, Y. Toyonaga, T.F.C. Mackay, M.D. Purugganan, and J. Schmitt. 2002. Novel loci control reproductive timing in *Arabidopsis thaliana* in natural environments. *Genetics* 162:1875-1884.
- Stinchcombe, J.R., C. Weinig, M. Ungerer, K.M. Olsen, C. Mays, S.S. Halldorsdottir, M.D. Purugganan, and J. Schmitt. 2004. A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *PNAS* 101: 4712-4717.
- Donohue, K. L. Dorn, C. Griffith, E. Kim, A. Aguilera, C. R. Polisetty, and J. Schmitt. 2005. Niche construction through germination cuing: life history responses to timing of germination in *Arabidopsis thaliana*. *Evolution* 59:771-785.
- Korves, T.M., K. J. Schmid, A. L. Caicedo, C. Mays, J.R. Stinchcombe, M.D. Purugganan, and J. Schmitt. 2007. Fitness effects associated with the major flowering time gene *FRIGIDA* in *Arabidopsis thaliana* in the field. *American Naturalist* 169:E141-E157.
- Wilczek, A., J. Roe, M. Knapp, M. Cooper, C.M. Lopez-Gallego, L. Martin, C. Muir, S. Sim, A. Walker, J. Anderson, J.F. Egan, R. Petipas, A. Giakountis, E. Charbit, G. Coupland, S.M. Welch, and J. Schmitt. Seasonal and geographic variation in sensitivity and balance of flowering pathways. *Science*, 15 Jan 2009. doi: 10.1126/science.1165826.

ii. Five additional publications:

- Weinig, C., J.R. Stinchcombe, and J. Schmitt 2003. QTL architecture of resistance and tolerance traits in *Arabidopsis thaliana* in natural environments. *Molecular Ecology* 12: 1153-1163.
- Donohue, K. L. Dorn, C. Griffith, E. Kim, A. Aguilera, C. R. Polisetty, and J. Schmitt. 2005. The evolutionary ecology of seed germination of *Arabidopsis thaliana*: variable natural selection on germination timing *Evolution* 59: 758-770.
- Mitchell-Olds, T., and J. Schmitt. 2006. Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis thaliana*. *Nature* 441:947-952.
- Von Wettberg, E.J., D.L. Remington, and J. Schmitt. 2008. Partitioning adaptive variation across a patchy landscape: shade avoidance traits in *Impatiens capensis*. *Evolution* 62: 654-667.
- Hopkins, R., J. Schmitt, and J.R. Stinchcombe. 2008. A latitudinal cline and response to vernalization in leaf angle and morphology in *Arabidopsis thaliana* (Brassicaceae). *New Phytologist* 179: 155-164.

D. Synergistic activities

i. Recent elected offices and honors:

President Elect (2007), President (2008), Past President (2009), SSE Council Delegate, Section on Biological Sciences (G), AAAS (2007-2010) President Elect (2001), President (2002), Past-President (2003-2005) ASN Alexander von Humboldt Research Award (2007) National Academy of Sciences (2008-)

ii. Selected recent professional service:

Faculty of 1000 (evolutionary ecology) 2004- NSF workshop on the future of Arabidopsis 2010 program, 2005. MASC Subcommittee on Natural Variation and Functional Genomics 2006- NSF workshop: 2020 Vision for Biology: the role of plants in addressing grand challenges 2008

iii Women in science:

Co PI: ADVANCE: Investing in leadership potential and faculty opportunities at Brown University (PI: P. O'Neil, co PIs B. Allen, M. Fennell, T. Palmore)

iv. Plant conservation:

Member, RI Task Force, New England Plant Conservation Program (NEPCoP) 1991- Member, NEPCoP Regional Advisory Committee 2000-

v. Undergraduate research mentoring:

48 Honors theses, over 30 publications with undergraduate coauthors

E. Collaborators and other affiliations

i. Collaborators last 48 months (undergraduate and postbac coauthors highlighted): A.

Aguilera (U Mass), C. Alonso-Blanco (CNB), R. Amasino (U. Wisc.), J. Anderson, J. Beder (U. Wisc. Milwaukee), E.W. Boyd (UICC), M. Blazquez (.Valencia), A. Caicedo (U. Mass), P. Carter (WSU), E.Charbit, G. Coupland (MPIZ), S. Das (KSU), C. Dean (John Innes), K. Donohue (Duke), L.A. Dorn (U. Wisconsin Oshkosh), ***J.F. Egan*** (Penn State), A. Giakontis (MPIZ); G. Gilchrist (William and Mary), R. Gomulkiewicz (WSU), C. Griffith, I. Gordon, ***K. Gravuer*** (Nature Serve), ***N. Hausmann*** (Berkeley), N. Heckman (UBC), M.S. Heschel (Colorado College), H. Huber (Radboud U.), J. G. Kingsolver (UNC), M. Kirkpatrick (U. Texas), ***J. Kelley*** (U.W.), E. Kim (Harvard), M. Knapp (KSU), T. Korves, M. Koornneef (MPIZ), C.M. Lopez-Gallego J. Marron (UNC), J. Martinez-Zapater (CNB), ***L. Martin*** (Cornell) ***L. McGeogh*** (U.C.Davis), C. Mays (NCSU), T. Mitchell-Olds (Duke), C. Muir (Indiana), M. Pigliucci (Stony Brook), K. Olsen (Washington U.), ***R. Petipas*** (U.VT), C. R. Polisetty (Kentucky), M. Purugganan (NYU), D. Remington (UNC Greensboro), ***C. Riginos*** (U.C. Davis); M. Reymond (MPIZ) J. Roe (KSU), O. Savolainen (Oulu), ***S. Sim*** (Notre Dame) J. Stinchcombe (Toronto), ***A. Walker*** (AAAS) D. Weigel (Salk/MPI); C. Weinig (Wyoming), S. Welch (KSU)

ii. Graduate and postdoctoral advisors:

Ph.D. advisor: Ward Watt, Stanford University
Postdoctoral advisor: Janis Antonovics, University of Virginia

iii. Ph.D. and postdoctoral advisees:

Ph.D. students (6 total): P. O'Neil (Assoc. Provost, Brown U.), C. Purrington (Assoc. Prof., Swarthmore), C.R. Linder (Assoc. Prof., U. Texas Austin), M.S. Heschel, (Asst. Prof, Colorado College); E. Von Wettberg (NIH postdoc fellow, UC Davis), Y. Toyonaga
Postdocs last 5 years (11 total): J. Stinchcombe (Assistant Professor, U. of Toronto), Tonia Korves, Bhoomika Singh, Amity Wilczek

NAME: Stephen M. Welch

PROFESSIONAL PREPARATION:

B.S., Computer Science, Michigan State University	Dec. 1971
Ph.D., Zoology, Michigan State University	Feb. 1977
Postdoctoral Fellow, Entomology, Michigan State University	Feb.-Oct. 1977

APPOINTMENTS:

1990-present	Full Professor, Department of Agronomy. Theoretical plant modeling. Acting State Climatologist (1989-92).
1993 (6 mo.)	Assistant to the Vice Provost for Academic Services. Assisted in general technology-related issues and activities as directed by the VPAS.
1992 (5 mo.)	Interim Director, KSU Computing and Network Services. Overall administration of networking, mainframe computing, public computing laboratories, and support activities for Kansas State University.
1987-1989	Computer Systems Coordinator, Kansas Cooperative Extension Service. Overall administration of the Computer Systems Office (which included the KSU Weather Data Library). Policy, planning, personnel, and budgetary management.
1985-1986	Technical Development Coordinator, Kansas Cooperative Extension Service. Needs assessment, financing, and administration of electronic technology development programs for information delivery.
1981-1984	Associate Professor, Department of Entomology, Kansas State University. Design of decision support systems for selected agricultural commodities.
1982 (3 mo.)	Visiting Professor, North Central Computer Institute, Madison, Wisconsin. Computerized teleconferencing applications in multi-state cooperative projects.
1977-1981	Assistant Professor, Department of Entomology, Kansas State University. Pest management modeling of economic insects.

PUBLICATIONS:

Related to proposal:

Wilczek A, Roe J, Knapp M, Cooper M, Lopez-Gallego CM, Martin L, Muir C, Sim S, Walker A, J.Anderson, Egan J.F., Petipas R., Giakountis A., Charbit E., Coupland G., Welch S.M., and Schmitt J. Seasonal and geographic variation in sensitivity and balance of flowering pathways. Science, 15 Jan 2009. doi: 10.1126/science.1165826.

Koduru P, Dong Z, Das S, Welch SM, Roe J, Charbit E. “Multi-Objective Evolutionary-Simplex Hybrid Approach for the Optimization of Differential Equation Models of Gene Networks”, IEEE Transactions on Evolutionary Computation, Vol. 12, No. 5, pp. 572 –590, 2008.

Hammer G, Cooper M, Tardieu F, Welch S, Walsh B, van Eeuwijk F, Chapman S, Podlich D. 2006. Models for navigating biological complexity in breeding improved crop plants. Trends in Plant Science 11:587-593.

Welch, S.M., J.L. Roe, S. Das, Z. Dong, R. He, M.B. Kirkham. 2005. Merging genomic control networks with soil-plant-atmosphere-continuum (SPAC) models. Agricultural Sys. 86:243-74.

Welch, S.M., Roe, J.L., Dong, Z. 2003. A genetic neural network model of flowering time control in Arabidopsis thaliana. Agronomy Journal, 95:71-81.

Other related publications:

Bulatewicz, T., Andresen, D., Welch, S.M., Jin, W., Das, S., Miller, M., 2007. A software system for scalable parameter estimation on clusters. In: Proc. of 8th LCI Int. Conf. on High-Performance Clustered Computing. <http://www.linuxclustersinstitute.org/conferences/archive/2007/technicalpapers.html>

Weinig C, Brock MT, Dechaine JA, Welch SM. 2006. Resolving the genetic basis of invasiveness and predicting invasions. Genetica. DOI 10.1007/s10709-006-9015-7.

Welch, S. M., G. Wilkerson, K. Whiting, N. Sun, T. Vagts, G. Buol, and T. Mavromatis. 2002. Estimating soybean model genetic coefficients from private sector variety performance trial

data. Transactions ASAE, 45(4):1163-75.
Chen, X., S. M. Welch, N. Zhang, D. Armbrust. 2002. Measurement of change in soybean plant cross-sectional area under wind conditions via image processing. Trans. ASAE, 44(6):1923-29.
Roman-Paoli, E., S.M. Welch, and R.L. Vanderlip. 2000. Comparing genetic coefficient estimation methods using the CERES-Maize model. Agricultural Systems, 65:29-41.

SYNERGISTIC ACTIVITIES:

- Dr. Welch is KSU co-PI in an NSF Frontiers of Integrative Biological Research (FIBR) project. The project involves four US institutions with unfunded collaborators at six European plant molecular genetics labs. Its goal is to study the molecular evolutionary ecology of developmental signaling pathways of *Arabidopsis thaliana* in complex environments from the DNA sequence to continental scales. The five-year project combines approaches from molecular biology, molecular evolutionary genetics, quantitative genetics, evolutionary ecology, gene network modeling, and biogeography. The KSU team is responsible for the mathematical modeling of genetic networks and the micro-meteorological monitoring at all field sites.
- Dr. Welch is a member of the Steering Committee for the Ecological Genomics Institute (EGI) in Kansas. EGI links responses of living systems to environmental change at the genetic level. Most environmental change studies span only one or two levels in the biological hierarchy, but biological responses depend on both genotype and gene-by-environment interactions, the relationships studied here. The EGI integrates the activities of 35 faculty and postdocs.
- Dr. Welch is a member of the Executive Committee of the Consortium for Global Research on Water-based Economies. GRoWE is developing coupled geo-spatial models for groundwater systems that include hydrological, plant, economic, and demographic components. It interacts closely with RIZA, the Dutch research and advisory body for inland waters management, renowned for its broad expertise in water issues world-wide. GRoWE also uses constructivist methods in educational experiments on teaching interdisciplinary topics in Water & Society.

COLLABORATORS & OTHER AFFILIATIONS:

Collaborators and Co-Editors: Richard Amasino, UW; Dean Armbrust, KSU; Kenneth Boote, UFL; Bruce Bugbee, USU; Scott Chapman, CSIRO; Xingwen Chen, Tetra-Tech, Inc. (Fairfax VA); Mark Cooper, PIO; Sanjoy Das, KSU; Jenny Dechaine, UMN; Andre de Roos, UAM; Zhanshan Dong, PIO; Kathleen Donohue, HU; Graeme Hammer, UQ; William Hsu, KSU; Ayse Irmak, UFL; Mary Beth Kirkham, KSU; Praveen Koduru, KSU; Jan Leach, CSU; James Jones, UFL; Tin Yu Mak, Alltel, Inc (Overland Park, KS); Theo Mavromatis, UFL; Jack Oviatt, KSU; Jeffrey Peterson, KSU; Dean Podlich, PIO; Michael Purugganan, NYU; Judith Roe, KSU; Elvin Roman-Paoli, UPR; Ruiqing He, UU; William Schapaugh, KSU; Johanna Schmitt, BU; David Steward, KSU; Naijian Sun, Microsoft Corp, (Seattle, WA); Francois Tardieu, IE; Harold Trick, KSU; Fred van Eeuwijk, WU; Richard Vanderlip, KSU; Bruce Walsh, UA, Virgil Valentine, KSU; Cynthia Weinig, UWY; Gail Wilkerson, NCSU; Naiqian Zhang, KSU.

Graduate Advisor: William Cooper, MISU **Postdoctoral Advisor:** Brian Croft, ORSU

Thesis Advisor to: M.S – Xingwen Chen Tetra-Tech, Inc. (Fairfax VA), Kai Ma KSU; PhD – Zhanshan Dong PIO, Praveen Koduru KSU, Xinye Cai KSU; Total advisory committee participation – MS 10, Ph D 14; Postdocs sponsored 3.

Affiliation codes: BU Brown University, CSIRO Commonwealth Scientific and Industrial Research Organisation, CSU Colorado St Univ, HU Harvard, IE INRA-ENSAM, KSU Kansas St Univ, NCSU North Carolina St Univ; NYU, New York, University; OHSU Ohio St Univ, ORSU Oregon St Univ, PIO Pioneer Hi-Bred International, SIS Space Imaging Services, UFL Univ of Florida, UGA Univ of Georgia, UA Univ Arizona, UAM Univ Amsterdam, UQ Univ Queensland, USU Utah St Univ, UU Univ Utah, UW Univ Wisconsin, WU Wageningen Univ, UWY University of Wyoming.

Curriculum Vitae

Name: Jeffrey W. White

Position:

Plant Physiologist, Plant Physiology and Genetics Group
U.S. Arid-Lands Agricultural Research Center, USDA, Agricultural Research Service
21881 N Cardon Lane, Maricopa, AZ 85238
phone: +1-520-316-6368; FAX: +1-520-316-6330 email: jwhite@uswcl.ars.ag.gov

A. Professional Preparation:

Harvard College, Cambridge, MA., Biology, B.A., 1975.
Univ. Calif., Berkeley, CA., Botany, Ph.D., 1981.

B. Appointments:

2003-present: Plant Physiologist; GS-15; U.S. Arid-Lands Agric. Res. Center
1995-2002: Head, GIS and Crop Modeling Laboratory; CIMMYT, Mexico
1983-1995: Bean Physiologist; Int. Center Tropical Agriculture (CIAT), Cali, Colombia
1981-1983: Post-doctoral fellow; International Potato Center (CIP), Lima, Peru

C. Recent Publications:

- White, J.W., M. Herndl, L.A. Hunt, T.S. Payne, and G. Hoogenboom. 2008. Simulation-based analysis of effects of *Vrn* and *Ppd* loci on flowering in wheat. *Crop Sci.* 48:678-687.
- White, J.W., van Evert, F.K., 2008. Publishing agronomic data. *Agron. J.* 100:1396-1400.
- White, J.W., Hoogenboom, G., Stackhouse Jr., P.W., and Hoell, J.M., 2008. Evaluation of NASA satellite- and assimilation model-derived long-term daily temperature data over the continental US. *Agric. Forest Meteor.* 148:1574-1584.
- White, J.W., K.J. Boote, G. Hoogenboom, and P.G. Jones. 2007. Regression-based evaluation of ecophysiological models. *Agron. J.* 99:419-427.
- White, J.W., 2006. From genome to wheat: Emerging opportunities for modelling wheat growth and development. *Eur. J. Agron.* 25:79-88.
- White, J.W., Hoogenboom, G., 2005. Integrated viewing and analysis of phenotypic, genotypic and environmental data with "GenPhEn arrays". *Eur. J. Agron.* 23, 170-182.
- Edmeades, G. O., McMaster, G. S., White, J. W., and Campos, H., 2004. Genomics and the physiologist: bridging the gap between genes and crop response. *Field Crops Res.* 90, 5-18.
- Hoogenboom, G., White, J.W., Messina, C.D., 2004. From genome to crop: integration through simulation modeling. *Field Crops Res.* 90, 145-163.
- Hoogenboom, G., and J. W. White. 2003. Improving physiological assumptions of simulation models by using gene-based approaches. *Agron. J.* 95:82-89.
- White, J.W. and G. Hoogenboom. 2003. Gene-based approaches to crop simulation: past experiences and future opportunities. *Agron. J.* 95:52-64.

D. Synergistic Activities:

1. The PI was co-leader of the International Crop Information System (ICIS) from 1997-2002, a collaborative effort of international centers, universities, and the private sector to develop databases and tools for documenting agricultural research. Since 2002, he has continued to promote use of ICIS, including helping establish ICIS databases for Lesquerella and Vernonia, two promising oil crops.
2. The PI is co-leader of the International Consortium for Agricultural Systems Applications (ICASA), which promotes use of simulation models and associated tools in agriculture.

Recently, this has included more effective use of models to simulate crop response to global change and novel strategies for including information from genomics in models.

3. The PI has explored new approaches for cross validation of simulation models, using multivariate analysis, mixed models, and quantile regression.

E. Collaborators & Other Affiliations

(ALARC = Arid Lands Agricultural Research Center, ret = retired or otherwise no longer directly affiliated)

Collaborators & Co-Editors

Ainsworth, EA, USDA-ARS, Urbana, IL

Boote, KJ, U of FL

Brooking, IR, N Z Inst. Crop & Food Res.

Bunce, JA, USDA-ARS, Beltsville, MD

Claupein, W, Univ. Hohenheim

Correll, M, Univ. Florida

Crossa, J, CIMMYT

DeLacy, IH, U Queensland

DelGrosso, S, USDA-ARS, Fort Collins, CO

Derner, JD, USDA-ARS, Cheyenne, WY

Dhillon, SS, Punjab Agri. Univ.

Dierig, D, ALARC

Edmeades, G, Pioneer (ret)

Fox, PN, ACIAR, Australia

Graeff, S, Univ. Hohenheim

Graham, P., U. Minnesota

Grant, RF, U Alberta

Greene, R, Virginia Tech

Hartkmap, AD, WAU (ret)

Hodson, D, CIMMYT

Hoell, J, NASA, Langley, VA

Hoogenboom, G, U of GA

Hunt, L.A., U Guelph

Jamieson, P. N Z Inst. Crop & Food Res.

Jones, JW, U of FL

Graduate Advisors and Postdoctoral Sponsors

Kaplan, DR (deceased), Botany, UC Berkeley (ret.)

Thesis Advisor and Postgraduate-Scholar Sponsor

Herndl, M. Univ. Hohenheim

Nath, D. Arizona State Univ.

(Total 10+ graduate students, 2 post doctoral fellows over career)

Jones, P, CIAT

Kimball, B USDA-ARS ALARC

McLaren, CG, IRRRI

McMaster, G, USDA-ARS, Ft. Collins, CO

Messina, C, Pioneer Int.

Montes, C, Univ. Popayan, Colombia

Morgan, J, USDA-ARS, Ft. Collins, CO

Ortiz-Monasterio, I., CIMMYT

Ottman, MJ, Univ AZ

Payne, T., CIMMYT

Porter, C, Univ. Florida, Gainesville, FL

Porter, J, K VL, Denmark

Ransom, J, NDSU

Reynolds, MP, CIMMYT

Salywon, A. Desert Bot. Garden, Phoenix

Sayre, KD, CIMMYT

Semenov, M, Rothamsted Res.

Shafer, SR, USDA-ARS, Peoria, IL

Sommer, R, ICARDA

Stackhouse, P, NASA, Langley, VA

Tomasi, P, ALARC

Trethowan, R, Univ. Sydney

van Evert, F, Plant Res. Int., The Netherlands

van Ittersum, M, WAU, The Netherlands

Wall, GW, ALARC

Sadik, S, International Potato Center

Rossi, A. Arizona State Univ.

Suggested Reviewers

Mark Cooper, Pioneer Hibred International, mark.cooper@pioneer.com
Graeme Hammer, Univ. Queensland, g.hammer@uq.edu.au
Bertram Ludäscher, UC Davis, ludaesch@ucdavis.edu
Zoé Lacroix, Arizona State University, Zoe.Lacroix@asu.edu
Holger Meinke, Wageningen University, holger.meinke@wur.nl
Gary Bader, University of Toronto, gary.bader@utoronto.ca
Ed Marcotte, University of Texas, marcotte@icmb.utexas.edu
John Kelly, University of Kansas, jkk@ku.edu
Tim Wheeler, University of Reading, t.r.wheeler@reading.ac.uk
James Kiniry, USDA/ARS Grassland Soil and Water Research Laboratory,
jim.kiniry@ars.usda.gov
Dennis Egli, University of Kentucky, degli@uky.edu
Matthijs Tollenaar, University of Guelph, mtollena@uoguelph.ca
Robert Grant, University of Alberta, robert.grant@ales.ualberta.ca
Doug Schemske, Michigan State University, schem@msu.edu
Toby Bradshaw, University of Washington, toby@u.washington.edu
Stephen Tonsor, University of Pittsburg, tonsor@pitt.edu